

West Virginia General Summative Assessment

2021–2022

**Volume 7
Special Studies**



West Virginia DEPARTMENT OF
EDUCATION

TABLE OF CONTENTS

- 1. Alignment Analysis of West Virginia General Summative Assessment for ELA and Mathematics Grades 3-8 with Corresponding Grade Level West Virginia College- and Career-Readiness Standards**
- 2. Device Comparability**
- 3. DIF Accommodations Analysis**
- 4. The Impact of the Pandemic on West Virginia Students**
- 5. Equating Test Scores across Administrations in the Context of the Covid-19 Pandemic**
- 6. Overall Results of an Alignment Analysis of the Cambium Statewide Science Assessment Item Bank for States Sharing Science Content with Corresponding Grade Level Next Generation Science Standards (NGSS) Performance Expectations (PEs)**
- 7. Alignment Analysis of the West Virginia General Summative Assessment (WVGSA) for Science Grades 5 and 8 with Corresponding West Virginia Next Generation Content Standards and Objectives for Science**
- 8. West Virginia 2022 COVID-19 Impact Study**

This volume consists of independent special studies conducted by CAI, AIR, WVDE, and WCEPS. This volume will be updated each year to reflect studies relevant to the respective administration.

REPORT

Alignment Analysis of West Virginia General Summative Assessment for ELA Grades 3-8 and Mathematics Grades 3-8 with Corresponding Grade Level West Virginia College- and Career-Readiness Standards

**Sara C. Christopherson and Norman L. Webb
February 27, 2019**

Wisconsin Center for Educational Products and Services
Matt Messinger, Executive Director
University of Wisconsin Research Park
510 Charmany Drive, Suite 269
Madison, WI 53719

Acknowledgements

English Language Arts Grades 3-5:

External Panelists

Anna Hardway	Group Leader	Texas
Chelsea Barnes		Wisconsin
Megan Lussig		Illinois

West Virginia Panelists

Stephanie Bruffy-Moffatt		West Virginia
Julie Craig		West Virginia
Kristen Hines		West Virginia

English Language Arts Grades 6-8:

External Panelists

Cindy Jacobson	Group Leader	Wisconsin
Kimberly Hernandez		Georgia
Tim Kreig		Indiana

West Virginia Panelists

Lara Parsons		West Virginia
Brooke Davis		West Virginia

Mathematics Grades 3-5:

External Panelists

Linda Hall	Group Leader	Washington, D.C.
Diane Briars		Pennsylvania
Denise Schultz		North Carolina

West Virginia Panelists

Tamatha Bock		West Virginia
Jillian Tieman		West Virginia
Desiree Fritsch		West Virginia

Mathematics Grades 6-8:

External Panelists

Lynn Raith	Group Leader	Pennsylvania
Diana Snider		Florida
Jaclyn Snyder		Pennsylvania

West Virginia Panelists

Stephanie Humienny		West Virginia
Teresa Atkinson		West Virginia
Ken Osborne		West Virginia

This independent third-party analysis was conducted for the West Virginia Department of Education (WVDE) and funded by the American Institutes for Research, Washington, D.C. Vaughn Rhudy, Executive Director of Assessment and Timothy Butcher, Lead Coordinator of Assessment, were the main contacts for West Virginia. Terri Sappington, Stacey Murrell, Sonya White, Robin Sizemore, Sonja Phillips, and Carrie Christy, all of WVDE, also attended the in-person study for the purpose of providing support as needed for questions specific to West Virginia (e.g. related to state standards or other state documentation).

Table of Contents

Table of Contents

Executive Summary	3
Introduction and Methodology	8
Training and Coding	10
Data Analysis	14
Alignment Criteria Used for This Analysis.....	16
Reporting Categories and Standards.....	16
Categorical Concurrence	17
Depth-of-Knowledge Consistency.....	18
Range-of-Knowledge Correspondence.....	18
Balance-of-Representation	19
Source of Challenge	19
Cutoffs for Alignment Criteria: Independent Analysis	20
Confirmatory Analysis Criterion	20
Cutoffs for Alignment Criteria: Confirmatory Analysis	20
Findings: Overall Program Capacity to Generate Aligned Assessments	21
Findings: ELA.....	25
Standards.....	26
Test Forms	27
Alignment Statistics and Findings	29
Results by Test Form: Independent Analysis.....	32
Results by Test Form: Confirmatory Analysis	45
Reliability among Reviewers.....	50
Findings: Mathematics	52
Standards.....	52
Test Forms	54
Alignment Statistics and Findings	55

Results by Test Form: Independent Analysis.....	57
Results by Test Form: Confirmatory Analysis	69
Reliability among Reviewers.....	72
Conclusion	73
References.....	76

For each content area:

Appendix A: Group Consensus DOK Values for West Virginia College- and Career-Readiness Standards

Appendix B: Data Analysis Tables for Each Test Form

Appendix C: Reviewers' Notes [Redacted for Public Release]

Appendix D: Debriefing Summary Notes [Redacted for Public Release]

Appendix E: DOK Definitions for Reading and for Mathematics

Appendix F: Coding Instructions Provided to Panelists

Executive Summary

This report describes an alignment analysis conducted during summer 2019, to provide information about the degree of alignment of the computer adaptive West Virginia statewide general summative assessments of Grades 3-8 ELA and Grades 3-8 Mathematics with the corresponding West Virginia College- and Career-Readiness Standards, as pertains to fulfilling requirements as stated in Federal statute.

The overall alignment analysis conducted for the West Virginia Department of Education incorporated information from assessment framework documents including blueprints and narrative information about the item selection algorithm, from statistical analyses provided within assessment technical reports of aggregate data from all administered test events in the state, and from a two-part content analysis. Consideration of framework and statistical documentation were conducted remotely. These components attended to the assessment design and the test developer's claims. The other study component involved a two-part content analysis. Content analyses were conducted by four (4) panels of six (6) expert educators during an in-person meeting. The overall study was designed to allow for the potential to craft a logic argument for the capacity of alignment of all test events generated by the West Virginia statewide general summative computer adaptive test (CAT) assessments for Grades 3-8 ELA and Grades 3-8 Mathematics with the corresponding West Virginia College- and Career-Readiness Standards, as appropriate, based on results.

A three-day in-person alignment institute took place on June 17 - 19, 2019, in South Charleston, West Virginia to analyze the agreement between the West Virginia standards and multiple simulated test events for each grade and subject. Three West Virginia educators and three external reviewers participated in each subject-area panel (ELA Grades 3-5, ELA Grades 6-8, Mathematics Grades 3-5, and Mathematics Grades 6-8). External panelists were selected because of their notable K-12 education experience and content expertise. West Virginia panelists were selected via a statewide protocol designed to identify appropriate educators for participation in committee, panel, and other work. The results of the in-person content analysis were compared with framework and statistical documentation to make an inference about the capacity of the overall test program to generate aligned test events for each subject and grade.

The central research focus for the in-person alignment analysis was to determine 1) the degree to which the simulated test events that were used in the analysis were aligned with the West Virginia College- and Career-Readiness Standards for ELA and Mathematics and 2) the accuracy and appropriateness of the vendor's internal metadata. The vendor's internal metadata specified the vendor's intended assessment target for each item, including the standard and DOK level of the item. Because the West Virginia ELA and Mathematics assessments are computer adaptive, the accuracy of the internal metadata is critical; appropriate selection of items for use in test events per test blueprints/algorithm is dependent on the internal metadata, particularly the coding for targeted standards.

ELA: Based on a consideration of the aggregate statistics of administered test events and assessment documentation, including test blueprints and narrative information about the item selection algorithm, along with the results of the content analyses, the West Virginia Statewide Assessment System demonstrates the capacity to generate acceptably or fully aligned test events for ELA grades 3-6. This finding is based on the following evidence:

- The blueprint designs for grades 3-6 support the potential for alignment with corresponding grade-level standards.
- The CAT algorithm takes into consideration appropriate alignment factors including item complexity and breadth of sampling, and has constraints to limit overemphasis of any particular standard.
- All administered test events met blueprint specifications, as reported in the assessment technical report.
- Based on the results of the independent analysis, all test forms analyzed for ELA grades 3-6 were considered acceptably aligned with the corresponding grade-level West Virginia ELA standards.
- Based on the results of the confirmatory analysis, all ELA test forms analyzed met the expectation that an independent review of item metadata by an expert educator panel finds at least 90% of all items to partially or exactly assess the corresponding internally assigned standards.

Panelists considered the vast majority of ELA items (between 71% and 100% on each test form) to exactly target a core component of the internally coded standard. Independent reviewers agreed, overall, with the internally coded standards but not with the internally coded DOKs. Although there was disagreement with the internally coded DOKs, the independent analysis of test forms found that all reporting categories for all of the 11 Grades 3-6 ELA test forms analyzed met the criterion of Depth-of-Knowledge Consistency.

For all sample test events analyzed for ELA Grades 7 and 8, each reporting category met most alignment criteria but forms were found to need slight adjustments to meet the minimum cutoffs for acceptable alignment. The main alignment issues for Grades 7 and 8 ELA test events was unmet or weak DOK Consistency for both reading reporting categories on all ELA test events analyzed for these grades. Averaged across forms, addressing unmet DOK consistency would require the revision or replacement of about seven items per test form. The targeted distribution of DOK as specified on the test blueprints does not parallel the distribution of DOK in the corresponding grade-level standards. This could be a factor in the unmet DOK Consistency for ELA Grades 7 and 8. In these cases, the test blueprint specifies around 25% of the items on any test form are expected to be DOK 3. In contrast, around 70% of Grades 7 and 8 standards were considered DOK 3. With some adjustments for DOK Consistency, the ELA assessment for Grades 7 and 8 would be expected to generate test events that are considered acceptably aligned with corresponding grade level standards.

All ELA test forms showed unmet or weak Range of Knowledge for the Writing and Language reporting category. It is important to note that this unmet Range is only in relation to the full set of standards specified by WVDE for use in the alignment study. This full set of standards includes three standards that are often omitted by states for the context of on-demand summative assessments: two that emphasize “guidance and support from peers and adults” (Standards 24 and 25) along with a focus on use of the Internet for purposes including interacting and collaborating with others (Standard 25) and one that includes an expectation for writing “routinely” (Standard 29) over the course of the year. States often omit these standards in the context of a one-time, on-demand assessment that is expected to be completed independently. It may also be appropriate to consider the three Text Types and Purposes writing standards (argumentative, informative/explanatory, and narrative) as a single mega-standard as, per test blueprint, WVDE does not expect students to complete more than one essay on the assessment. If the Writing and Language reporting category were viewed as such, it would include 10-11 standards per grade instead of 15-16, resulting in more test events meeting the Range-of-Knowledge criterion. Regardless, unmet Range was not considered a significant alignment issue for any test event analyzed. When this was the only alignment gap, as for Grades 3-6 ELA, the overall test events were still considered acceptably aligned (defined as needing 5 or fewer items revised or replaced to fully meet minimum alignment criteria) with the corresponding grade level standards because the issue could be resolved with the revision or replacement of just a few items for each test form.

All writing prompts for Grades 3-8 were considered to target multiple standards within the Writing and Language reporting category. Based on the three-part rubric, student responses were evaluated in relation to both Writing standards and Language standards within the reporting category.

Mathematics: Based on a consideration of the aggregate statistics of administered test events and assessment documentation, including test blueprints and narrative information about the item selection algorithm, along with the results of the content analyses, the West Virginia Statewide Assessment System demonstrates the capacity to generate acceptably or fully aligned test events for Mathematics Grades 3-8. This finding is based on the following evidence:

- The blueprint designs support the potential for alignment with corresponding grade-level standards.
- The CAT algorithm takes into consideration appropriate alignment factors including item complexity and breadth of sampling, and has constraints to limit overemphasis of any particular standard.
- All administered test events met blueprint specifications, as reported in the assessment technical report.
- Based on the results of the independent analysis, all Mathematics test forms analyzed were considered fully aligned or acceptably aligned with the corresponding grade-level West Virginia Mathematics standards.
- Based on the results of the confirmatory analysis, all Mathematics test forms analyzed met the expectation that an independent review of item metadata by an expert educator panel finds at least 90% of all items to partially or exactly assess the corresponding internally assigned standards.

Panelists considered the vast majority of Mathematics items (between 76% and 100% on each test form) to exactly target a core component of the internally coded standard. Despite some disagreement with internally coded DOKs, only one reporting category on one test form did not meet DOK consistency—resolved with the revision or replacement of just two items. Differences in reviewer and vendor DOK assignments for Mathematics items may reflect differences in training, in use of references for the conceptual model of complexity, or other factors, but are not considered an alignment issue.

Overall, the results of a two-part content alignment analysis along with a consideration of blueprints, information about the item selection algorithm, and aggregate data from administered test events as reported in the assessment technical report suggest that the West Virginia statewide general summative computer adaptive test program for Grades 3-8 Mathematics and Grades 3-6 ELA has the capacity to generate aligned test events across a range of proficiency. The evidence to support this claim includes:

- A confirmatory content analysis of items on each of 11 test events for ELA Grades 3-6 and 19 test events for Mathematics Grades 3-8 suggests that the items are appropriate as relates to intended claims and inferences about the targeted content (standard) based on comparison of an internal metadata with independent external analysis.
- According to the aggregate data for all administered test events as shown in the 2018-2019 West Virginia Technical Report, the test blueprints and algorithm are generating test events as intended; with the exception of known and uncommon circumstances, all tests delivered met blueprint specifications.
- Further, an independent analysis of test events sampled from a range of proficiency levels for Grades 3-8 Mathematics and Grades 3-6 ELA were found to fully or acceptably meet minimum alignment criteria using typically accepted cutoffs for Categorical Concurrence, DOK Consistency, Range of Knowledge Correspondence, and Balance of Representation.

For all test events analyzed for ELA Grades 7 and 8, all reporting categories met most alignment criteria but all test events were found to need adjustments to meet the minimum cutoffs for overall alignment. To be fully aligned, each test event required the revision or replacement of around 7 items per test form to address weak or unmet DOK consistency for both reading reporting categories. Revisions or amendments to the Grade 7 and 8 item bank could ensure that there are a sufficient proportion of DOK 3 items corresponding to DOK 3 standards. WVDE may also wish to review the specifications for distribution of DOK (“DOK Ranges”) given in the test blueprints in relation to the distribution of DOK in the corresponding grade-level standards. The test blueprint specifies around 25% of the items on any ELA test form are expected to be DOK 3. In contrast, around 70% of Grades 7 and 8 standards were considered DOK 3. As noted previously, the external reviewers did not always agree with the internally coded DOK, so other factors could also be in play such as differences in training or in the use of references (Webb-sourced vs alternate interpretations). Because DOK consistency for ELA Grades 7 and 8 was the only significant alignment gap, the assessment program could be expected to generate acceptably aligned test events for these grades with appropriate revisions in place.

WVDE may also wish to review the blueprint specifications related to targeting Mathematical Habits of Mind (MHM) as panelists found very few items they thought truly elicited the intent of the MHMs in the summative assessment items. This is not considered an alignment issue, as MHMs are not necessarily appropriately measured in a single snapshot, as on a summative assessment. However, WVDE may wish to reconsider the claims about MHM sub-score reporting. For example, although the assessments do not elicit the full intent of the MHMs, the items may contribute to student engagement with some aspects of these MHMs.

Introduction and Methodology for Content Alignment Analyses

The alignment of expectations for student learning with assessments for measuring students' attainment of these expectations is an essential attribute for an effective standards-based education system. Alignment is defined as the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide an education system toward students learning what they are expected to know and do. As such, alignment is a quality of the relationship between expectations and assessments and not an attribute solely of either of these two system components. Alignment describes the match between expectations and an assessment that can be legitimately improved by changing either student expectations or the assessments. As a relationship between two or more system components, alignment is determined by using the multiple criteria described in detail in a National Institute for Science Education (NISE) research monograph, *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education* (Webb, 1997). The corresponding methodology used to evaluate alignment has been refined and improved over the last 20 years, yielding a flexible, effective, and efficient analytical approach, including modifications appropriate for analysis of multidimensional standards/assessments and computer-adaptive test programs.

The alignment analysis of West Virginia statewide general summative computer adaptive assessments of Grades 3-8 ELA and Grades 3-8 Mathematics with the corresponding grade level/grade band West Virginia College- and Career-Readiness Standards incorporated information from assessment framework documents including blueprints and narrative information about the item selection algorithm, from statistical analyses provided within assessment technical reports of aggregate data from all administered test events in the state, and from a two-part content analysis. Each study component addressed one of the four research questions listed below.

1. **Framework Analysis:** Do the CAT item selection algorithms, test blueprints, and other relevant test specifications and documentation reflect appropriate design to support potential alignment of test events with corresponding grade-level academic standards?
2. **Statistical Analysis:** Do the available aggregate data for recently administered test events provide evidence that the algorithm and blueprints are yielding test forms as expected?
3. **Content Analysis (Metadata):** Are the internally (AIR-)assigned standards for each item reasonably consistent with the codings of an independent panel of expert educators?
4. **Content Analysis (Test Events):** To what degree are actual test events aligned with corresponding grade-level academic standards?

Consideration of framework and statistical documentation were conducted remotely. The two-part in-person content alignment analysis of the computer adaptive West Virginia statewide general summative assessments of Grades 3-8 ELA and Grades 3-8 Mathematics with the corresponding West Virginia College- and Career-Readiness Standards was conducted during June 17 - 19, 2019, in South Charleston, West Virginia. The study was designed to yield information that could be used to judge 1) the degree to which the simulated test events that were used in the analysis were aligned with the West Virginia College- and Career-Readiness Standards for ELA and Mathematics and 2) the accuracy and appropriateness of the vendor's internal metadata. The vendor's internal metadata specifies the vendor's intended assessment target for each item, including the standard and DOK level of the item. A content analysis of items and the corresponding internal metadata was required to generate data that could be used as evidence to (potentially) substantiate the test developer's claims about what each item measures. Because the West Virginia ELA and Mathematics assessments are computer adaptive, the accuracy of the internal metadata is critical; appropriate selection of items for use in test events per test blueprints/algorithm is dependent on the internal metadata, particularly the coding for standards.

In their first round of coding for each test event, panelists independently assigned one or more standards and a level of complexity (DOK) to each item. These data were used to determine the degree to which the simulated test events analyzed addressed the full depth and breadth of the state standards for Grades 3-8 ELA and Grades 3-8 Mathematics. On their second round of coding for each test event, panelists completed a confirmatory analysis of the internal metadata. They reviewed the vendor-based coding for each item, rating the degree to which the item measured the internally assigned standard (*exactly, partially, minimally, or not at all*). Definitions were provided for each rating along with training for consistent use.

The test vendor provided four simulated test events for each grade and subject area. First, a pool of 1,000 simulated test events were generated across the full ability scale. Then, a test event was randomly selected from the cases close to or at a cut score. The West Virginia assessment program for ELA and Mathematics uses four achievement levels: (1) Does Not Meet Standard, (2) Partially Meets Standard, (3) Meets Standard, and (4) Exceeds Standard. One test event was sampled from Does Not Meet Standard (1), at or near the Partially Meets Standard (2) cut score. Two test events were sampled from Meets Standard (3), at or near the Meets Standard cut score. One test event was sampled from Exceeds Standard (4), at or near the Exceeds Standard cut score.

Sara Christopherson led the study in collaboration with Norman Webb who developed the core alignment study procedures and criteria (through the National Institute for Science Education in 1997, funded by the National Science Foundation, and in cooperation with the Council of Chief State School Officers) that influenced the specification of alignment criteria by the U.S. Department of Education. Adjustments to a fixed-form alignment approach are necessary for a computer adaptive alignment process (e.g. Wise, Kingsbury, and Webb, 2015). A modified Webb alignment process was therefore used for this study with adjustments appropriate for a computer adaptive context. Original and modified versions of the Webb alignment process have been used to analyze curriculum standards and assessments in at least 30 states to satisfy or to prepare to satisfy Title I compliance as required by the United States Department of

Education (USED). Sara Christopherson has participated in and led Webb alignment studies since 2005 for state departments of education as well as for other entities. An experienced group leader facilitated each panel.

The Version 2 of the Web Alignment Tool (WATv2) was used to enter all of the content analysis codes during the institute. The WATv2 is a web-based tool connected to the server at the Wisconsin Center for Education Research (WCER). It was designed to be used with the Webb process for analyzing the alignment between assessments and standards. Prior to the institute, a group was registered on the WATv2 for each of the four panels. Each panel was assigned one or more group identification numbers and the group leader was designated. Then the reporting categories and standards were entered into the WATv2 along with the information for each assessment, including the number of items/tasks, the weight (point value) given to each item/task, and additional comments such as the identification number for the item to help panelists find the correct item/task. Separate studies were set up in the WATv2 for panelists to record data for the confirmatory analysis.

Training and Coding

In the morning of the first day of the alignment institute, all reviewers received an overview of the purpose of their work, the coding process, and general training on the Depth-of-Knowledge (DOK) definitions used to describe content complexity. All reviewers reported some experience with the DOK levels prior to the institute. The general training at the alignment institute was crafted to contextualize the origins of DOK (to inform alignment studies of standards and assessments) and purpose (to differentiate between and among degrees of complexity), and to highlight common misinterpretations and misconceptions to help reviewers better understand and, therefore, consistently apply the Depth-of-Knowledge (DOK) language system. Panelists also practiced assigning DOK to sample assessment items that were selected to foster important discussions that promote improved conceptual understanding of DOK. Appropriate training of the panelists at the alignment institute is critical to the success of the project. A necessary outcome of training is for panelists to have a common, calibrated understanding of the DOK language system for describing categories of complexity. Panelists were also provided examples and opportunities to practice differentiating between and among the categories used for the confirmatory analysis. Discussion about the examples helped panelists calibrate their understanding and clarified the uses of the categories. Coding instructions and descriptions for each category used in the confirmatory analysis are included in **Appendix F**. The introductory training also provided an overview of how to navigate within the online tool used to record data (the WATv2) as well as the online platform used to view the test events (the vendor-based AIR Content Rater).

The Mathematics and ELA groups were then separated into different rooms to receive more detailed training on the DOK levels for each subject area. Through interactive and participatory training, panelists reviewed the subject area-specific definitions of the four DOK levels and worked toward a common understanding of the difference between and among each of the levels of complexity. Definitions for each DOK level for ELA and Mathematics are included within **Appendix E**. Reviewers then worked to calibrate their use of DOK as they evaluated the complexity of a subset of the standards, first assigning DOK individually and then participating in a consensus discussion. After completing

coding and discussion of the subset, the panelists completed assigning DOK levels to the standards. For efficiency, panelists were able to review previously assigned DOKs for most standards, flagging any that they thought needed to be reconsidered. Group leaders then facilitated consensus discussions for any standards for which panelists disagreed about the complexity. The standards analysis is a necessary component of the alignment study but also, importantly, fosters thorough, nuanced, and calibrated understanding of the standards by panelists. Consensus DOK levels were then entered into the online data collection system, the WATv2. The consensus DOK values for all standards are given in **Appendix A** for each subject. Additional detail about the subject-area standards discussions are provided within this report.

After thoroughly discussing the standards and coming to consensus on the intended complexity of each standard, both grade-band panels for each subject evaluated the same Grade 6 test event to check for inter-panel consistency in coding. Panelists coded the first 3-5 items on the Grade 6 test event and then adjudicated to ensure that all panelists understood the processes. For each item, panelists worked individually to assign a DOK level to the item and then to code each item to the standard that they judged the item measured, i.e. what students must demonstrate that they know or are able to do in order to successfully respond to the question. Up to three standards could be coded as corresponding to each item. Following individual analyses of the items, reviewers participated in a debriefing discussion in which they analyzed the degree to which they had coded particular items or types of content to the standards. This was another important opportunity for panelists to discuss differences in their coding of items to standards from a particular framework and sort out any confusion or misunderstandings at the very start. Any differences were discussed and group leaders facilitated group consensus to clarify standard interpretations, procedural aspects, and any other topics as needed. Panelists also practiced conducting the confirmatory analysis of internal codings and sorted out any differences in interpretation. For example, panelists needed to understand that they were considering the relationship between the internal metadata and the item versus the relationship between their own coding and the internal metadata. Panelists then continued work, independently reviewing each assessment item on a simulated test event followed by conducting a confirmatory analysis of the internally assigned standards for each test event.

To maintain calibration within each group of reviewers, grade-band panels completed the first 3-5 items and then debriefed for the first test event for each grade. This process was conducted for both the independent round of coding and the confirmatory round of coding. Reviewers then completed analysis of the remaining items individually for each test form and for each round of coding.

Reviewers were instructed to focus primarily on the alignment between the West Virginia standards and the assessment items on the test forms. However, reviewers were encouraged to offer their opinions on the standards and on the assessment tasks by writing a note about the item in the appropriate text box in the WATv2 data collection tool. Reviewers were instructed to enter a note into the WATv2 for an assessment item if the item only corresponded to a part of a standard and not the full standard. Thus, the reviewers' notes can be used to reveal if assessment items only targeted a part of the individual standards. Reviewers also could indicate whether there was a Source-of-Challenge issue with an item—i.e. a technical or content problem with the item that

might cause a student who knows the material to give a wrong answer or enable someone who does not have the knowledge being tested to answer the item correctly. No Source-of-Challenge issues were identified by more than one reviewer for any of the ELA test events. Two Mathematics items were flagged with Source-of-Challenge. Item-level reviewer comments were provided in full to the state and to the test vendor but have been redacted from this report for public release.

After completing coding of a test form, reviewers engaged in adjudication of their results for the first few test forms. After the first few test events, broad differences in coding were very limited and panels conducted very little or no adjudication. For the first few forms, after all reviewers completed coding the test event, the study director and group leader identified the assessment items that did not have a majority of reviewers in agreement on DOK or where the reviewers differed significantly on the DOK assigned (e.g. three different DOK values were assigned). When these substantial differences in DOK occur, it sometimes indicates a data entry error. If data are entered as intended, then it suggests that reviewers are either interpreting the DOK definitions in very different ways or are interpreting the particular assessment item in very different ways. Adjudication can help panelists calibrate their interpretation of these study components.

Reviewers also discussed items for which there were great differences in coding to a standard. The adjudication process helped panelists identify and correct any errors in coding (e.g. accidentally assigning an item to a standard that they did not intend to assign). Adjudication also helped panelists build familiarity with the standards (e.g. a reviewer might not have noticed that a particular expectation is explicit in one of the standards) as well as build common interpretation of the standards (e.g. panelists may calibrate their understanding of the meaning of certain standards that may be interpreted in different ways because of ambiguous wording or differences in the way people understand the content). Adjudication additionally helped reveal differences in interpretation of assessment items and helped reviewers to build a common understanding of exactly what knowledge, abilities, and skills particular items were assessing. Overall, adjudication is intended to foster full and appropriate interpretation of the assessment items and standards, and to ensure that panelists have coded the items as they intended. Reviewers were not required to change their results after the discussion. Reviewer agreement statistics were computed after adjudication and are included in the Findings section of this report.

Reviewers were instructed to consider the full statement of expectations to consider if an assessment item should be mapped to a standard. In some cases, reviewers could make reasonable arguments for coding an item to different standards. For example, ELA reviewers discussed how to differentiate between the multiple standards that include the expectation for students to use context to interpret the meaning of an unknown word or phrase (standards 7, 10, 39, and 40). Reviewers agreed that they could differentiate between and among the standards based on whether the word/phrase was presented in a Literary Reading context (standard 7), an Informational Reading context (standard 10) or a stand-alone type of context (standard 39 or 40, depending on content). Input from West Virginia educators and clarification from WVDE staff about state interpretation of the differences between and among these standards helped the group to decide on decision rules for assigning these standards to items. Although decision rules were

defined and agreed upon, panelists commented in their notes that they continued to struggle with differentiating between these standards for coding purposes.

For the confirmatory analysis, panelists were able to agree with the internal metadata even if they had selected a different standard independently. For example, and using the context described above, a panelist might have independently coded an item to ELA standard 7 but still agree that the item reasonably targeted the internally coded ELA standard 39.

If reviewers map an item to a variety of standards, it may indicate that the assessment task is not a close fit for any standard. If no particular grade-level standard was targeted by a given assessment item, then the reviewers were instructed to code the item to a standard where there was a partial, but reasonable, fit or to a conceptual category level: the domain or reporting category level. Coding to the level of a conceptual category is referred to as coding to a “generic” standard. This coding to a “generic standard” sometimes indicates that an item was inappropriate for a particular grade level (for example, the item might better match a standard from another grade level). If an item was grade-appropriate but a corresponding standard was not found, a generic coding may indicate that the item is targeting knowledge within the standards that is being interpreted differently by different parties. Across all 37 test events analyzed, only three ELA items and two Mathematics items were coded to generic standards. This indicates that for all test forms analyzed, reviewers found one or more specific state standard(s) that nearly every item targeted.

During the second day of the study (the first full day of coding), it was brought to the attention of the Study Director and the WVDE stakeholders that one ELA panelist was not following necessary processes despite multiple interventions. This panelist was observed not using the necessary references to input data (West Virginia standards packet, DOK definitions, assessment item, and data collection tool), and not following agreed upon procedures (e.g. how to enter data for the writing prompt, to reflect the three-part rubric). This reviewer also self-reported not understanding how to use the computers and that the data that the individual had already entered were not entered in the correct order. After discussion with WVDE stakeholders, it was decided that this panelist’s data would be stripped from the final data set but that the participation on the panel could continue as a professional development opportunity. After study completion, the full sets of data were downloaded. Then, this reviewer’s data were removed from all of the studies for which the reviewer had entered data. The data removal process was conducted by Robert Glover, Director of Technical Services at the Wisconsin Center for Education Research (WCER) at the University of Wisconsin-Madison. The dataset was then downloaded again, with the errant data removed. The results reported in this study do not include this reviewer’s data.

Each panelist used two laptops, provided by the test vendor, for the item-level coding work during the in-person alignment analysis. One laptop was used to view the items on each test event through the AIR Content Viewer and the other laptop was used for data entry into the online data collection system, the WATv2. The on-site IT support was very responsive to Internet as well as other computer issues. Reviewers were periodically kicked out of the AIR Content Viewer portal (item access); one group found that item IDs would switch mid-process but found they could reset them to the original ID by logging

off and on again to Content Viewer. One test form was not in the Content Viewer system. At the beginning of the data entry, a WATv2 (data entry tool) server “bottleneck” caused a slow-down of response. The WCER IT team responded and resolved the issue but some reviewer data were lost and had to be reentered. This issue took about one hour to resolve. Reviewers were able to continue analyzing items, recording data on paper, but there was some lost time because of the issues with data entry.

It was expected that most panelists could get through three test forms per grade and that at least a subset of reviewers could code four forms per grade. Of the four test forms, one was sampled from below proficiency, marked by grade and labeled as Batch 1. For example, the G3B1 test form was the below proficiency (Batch 1) test form for Grade 3. Test forms representative of students judged to be at proficiency were labeled as Batches 2 and 3 (e.g. G3B2 and G3B3) and the test forms sampled from above proficiency were labeled as Batch 4 (e.g. G3B4). Panelists were not provided information to decipher this coding to avoid any influence of perception of the below, at, or above proficiency designation for test forms. If needed due to the pace of group coding, goals were scaled back to include at minimum an analysis of one below-proficiency test event and one at-proficiency test event. The rationale for this decision was to prioritize the most vulnerable decision points. If panelists completed forms from Batches 1 and 2, they were instructed to then move to batch 4 (an above-proficiency test event) before moving on to other test forms. However, some panels completed two at-proficiency forms for a grade instead of moving on to the above-proficiency sample. The overall goal was to collect data across the broadest range of proficiency possible, as time allowed.

Panelists work through test forms at different rates, due to a variety of factors. If some panelists finished coding before others, they moved on to subsequent test forms. Several test forms were coded by a sub-group of three or more panelists. When this occurred, the group was required to be composed of both internal and external reviewers. Both a 1:2 or 2:1 ratio (at minimum) was acceptable. No three-person panel was composed of entirely West Virginia-based reviewers or of entirely external reviewers.

Data Analysis

To derive the results from the independent analysis, the reviewers’ responses were averaged. First, the value for each of the four alignment criteria (described in the next section) was computed for each individual reviewer. Then the final reported value for each criterion was found by averaging the values across all reviewers. Any variance among reviewers was considered legitimate, for example, with the reported DOK level for an item falling somewhere between the two or more assigned values. Such variation could signify differences in interpretation of an item or of the assessed content and/or a DOK that falls in between two of the four defined levels. Any large variations among reviewers in the final results represented true differences in opinion among the reviewers and were not because of coding error. These differences could be because of different standards targeting the same content knowledge or may be because an item did not explicitly correspond to any standard, but it could be inferred to relate to more than one standard. Standard deviations are reported in the tables provided in **Appendix B**, which give one indication of the variance among reviewers.

To derive the results from the confirmatory analysis, items were grouped and reported by category:

- majority of reviewers labeling the item as an **exact match** with the internally coded standard
- majority of reviewers labeling the item as an **exact or partial match** with the internally coded standard

Any items for which there was no majority or for which the majority of reviewers labeled **minimal or no match** are reported individually, by item ID, within the results. These items should be reviewed as they may be internally miscoded or else need to be adjusted or replaced.

The results for each content area produced from the institute pertain only to the issue of alignment between the West Virginia standards and the simulated test forms that were analyzed. Note that an alignment analysis of this nature does not serve as external verification of the general quality of the standards or assessments. Rather, only the degree of alignment is discussed in the results. For these results, the means of the reviewers' independent coding were used to determine whether the alignment criteria of Categorical Concurrence, DOK Consistency, Range of Knowledge, and Balance of Representation were met.

Alignment Criteria Used for This Analysis

This report describes the results of a two-part content alignment analysis of the computer adaptive West Virginia statewide general summative assessments of Grades 3-8 ELA and Grades 3-8 Mathematics with the corresponding West Virginia College- and Career-Readiness Standards. Data from the independent content analysis of test forms by expert educator panels were used to address specific criteria related to the agreement between the expectations within the standards and the demands of the items within the assessments. For both the ELA and Mathematics studies, four criteria received major attention:

- Categorical Concurrence,
- Depth-of-Knowledge Consistency,
- Range-of-Knowledge Correspondence, and
- Balance-of-Representation.

Details on the criteria and indices used for determining the degree of alignment between standards and assessments are provided below. For each alignment criterion, an acceptable level was defined by what would be required to ensure that a student had reasonably met the expectations within the reporting categories for each discipline. In the descriptions below, the words “domain” and “reporting category” are used to describe reporting levels.

Reporting Categories and Standards:

Study results are reported according to the reporting categories (RCs) for each content area. These RCs are given below. Consensus DOK values for all standards are given in **Appendix A** for each subject.

In this analysis, the reporting categories for **Grades 3-8 ELA** were:

- Literary Reading (LR)
- Informational Reading (IR)
- Writing and Language (WL)

Total number of standards: 32-33

The reporting categories for **Grades 3-5 Mathematics** were:

- Operations and Algebraic Thinking (OA)
- Number and Operations in Base Ten and Fractions (NBT.NF)
- Measurement and Data and Geometry (MD.G)

Total number of standards: 34 (Grade 3); 36 (Grade 4); 34 (Grade 5)

The reporting categories for **Grade 6 Mathematics** were:

- Ratios and Proportional Relationships and The Number System (RP.NS)
- Expressions and Equations (EE)
- Geometry and Statistics and Probability (G.SP)

Total number of standards: 37 (Grade 6)

The reporting categories for **Grade 7 Mathematics** were:

- Ratios and Proportional Relationships and The Number System (RP.NS)
- Expressions and Equations (EE)
- Geometry (G)
- Statistics and Probability (SP)

Total number of standards: 34 (Grade 7)

The reporting categories for **Grade 8 Mathematics** were:

- The Number System and Expressions and Equations (NS.EE)
- Functions (F)
- Geometry and Statistics and Probability (G.SP)

Total number of standards: 36 (Grade 8)

In the descriptions below, the term “standards” may be used as an umbrella term, to refer to expectations in general. In addition to judging alignment between reporting categories and assessments on the basis of the four key alignment criteria, reviewers had the opportunity to comment on all items.

Categorical Concurrence

An important aspect of alignment between standards and assessments is whether both address the same content categories. The Categorical-Concurrence criterion provides a very general indication of alignment if both documents incorporate the same content.

The criterion of Categorical Concurrence between standard and assessments is met if the same or consistent categories of content appear in both documents. This criterion was judged by determining whether the assessment included items measuring content from each reporting category. The analysis assumed that the assessment had to have at least six items (or points for polytomous items) for measuring content from a reporting category for a minimum acceptable level of Categorical Concurrence to exist between the domain and the assessment. The number of items/points, six, is based on estimating the number of items that could produce a reasonably reliable subscale for estimating students’ mastery of content on that subscale. Of course, many factors must be considered in determining what a reasonable number is, including the reliability of the subscale, the mean score, and cutoff score for determining mastery. Using a procedure developed by Subkoviak (1988) and assuming that the cutoff score is the mean and that the reliability of one item is 0.1, it was estimated that six items would produce an agreement coefficient of at least 0.63. This indicates that about 63% of the group would be consistently determined to be masters or non-masters if two equivalent test administrations were employed. The agreement coefficient would increase to 0.77 if the cutoff score is increased to one standard deviation from the mean and, with a cutoff score of 1.5 standard deviations from the mean, to 0.90.

Usually states do not report student results by domains or require students to achieve a specified cutoff score on expectations related to a domain. If a state did do this, then the state would seek a higher agreement coefficient than 0.63. Six items were assumed as a minimum for an assessment measuring content knowledge related to a reporting category, and as a basis for making some decisions about students’ knowledge of that content under the reporting category. If the mean for six items is 3.0 points and one standard deviation is equal to a one-point item, then a cutoff score set at 4.0 points

would produce an agreement coefficient of 0.77. Any fewer items with a mean of one-half of the items would require a cutoff that would only allow a student to miss one item. This would be a very stringent requirement, considering a reasonable standard error of measurement on the subscale.

Depth-of-Knowledge Consistency

Standards and assessments can be aligned not only on the category of content covered by each, but also on the basis of the complexity of knowledge required by each; i.e. how students are expected to cognitively engage with the content. *Depth-of-Knowledge Consistency between standards and an assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards.* For consistency to exist between the assessment and the reporting categories, as judged in this analysis, at least 50% of the items corresponding to a reporting category had to be at or above the depth-of-knowledge level of the corresponding content expectation. The 50% level, a conservative minimum cutoff point, is based on the assumption that a minimal passing score for any one reporting category of 50% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the content expectations within the corresponding reporting categories. For example, assume an assessment included six items related to one domain and students were required to answer correctly four of those items to be judged proficient—i.e. 67% of the items. If three, 50%, of the six items were at or above the depth-of-knowledge level of the corresponding expectations, then for a student to achieve a proficient score would require the student to answer correctly at least one item at or above the depth-of-knowledge level of one expectation. If a domain had between 40% and 50% of items at or above the depth-of-knowledge levels of the expectations, then it was reported that the criterion was “weakly” met.

DOK Levels 1-4

Interpreting and assigning depth-of-knowledge levels to both standards and assessment items is an essential requirement of alignment analysis. The DOK descriptions help to clarify what the different levels represent for Reading and Mathematics. Full descriptions for each subject are included in **Appendix E**.

Range-of-Knowledge Correspondence

For reporting categories and assessments to be aligned, the breadth of knowledge required on both should be comparable. *The Range-of-Knowledge criterion is used to judge whether a comparable span of knowledge expected of students by a reporting category is the same as, or corresponds to, the span of knowledge that students need to correctly answer the assessment items/activities.* The criterion for correspondence between the span of knowledge for a reporting category and an assessment considers the number of standards within the reporting category with one related assessment item/activity. Fifty percent of the standards for a reporting category must have at least one related assessment item for the alignment on this criterion to be judged acceptable. This level is based on the assumption that students’ knowledge should be tested on content from over half of the domain of knowledge for a reporting category. This assumes that each expectation for a reporting category should be given equal weight. Depending on the balance in the distribution of items and the need to have a low number of items related to any one expectation, the requirement that assessment items need to

be related to more than 50% of the expectations for a reporting category increases the likelihood that students will have to demonstrate knowledge on more than one expectation per reporting category to achieve a minimal passing score. As with the other criteria, a state may choose to make the acceptable level on this criterion more rigorous by requiring an assessment to include items related to a greater number of the expectations. However, any restriction on the number of items included on the test will place an upper limit on the number of expectations that can be assessed. Range-of-Knowledge correspondence is more difficult to attain if the content expectations are partitioned among a greater number of reporting categories and if there are a large number of expectations. If 50% or more of the objectives for a reporting category had a corresponding assessment item, then the Range-of-Knowledge correspondence criterion was met. If between 40% and 50% of the objectives for a reporting category had a corresponding assessment item, the criterion was “weakly” met.

Balance-of-Representation

In addition to comparable depth and breadth of knowledge, aligned reporting categories and assessments require that knowledge be distributed equally or proportionally in both. The Range-of-Knowledge criterion only considers the number of expectations with at least one assessment item within a reporting category; it does not take into consideration how the assessment items/activities are distributed among these expectations. *The Balance-of-Representation criterion is used to indicate the degree to which one standard is given more emphasis on the assessment than another.* An index is used to judge the distribution of assessment items. This index only considers the expectations for a reporting category that has at least one related assessment item for one or more expectations. The index is computed by considering the difference in the proportion of expectations and the proportion of items assigned to the expectation. An index value of 1 signifies perfect balance and is obtained if the corresponding items related to a reporting category are equally distributed among the expectations for the given reporting category. Index values that approach 0.0 signify that a large proportion of the items assess only one or two of all of the expectations that were measured. Depending on the number of expectations and the number of items, a unimodal distribution (most items related to one expectation and only one item related to each of the remaining expectations) has an index value of less than 0.5. A bimodal distribution has an index value of around 0.55 or 0.6. Index values of 0.7 or higher indicate that items/activities are distributed among all of the expectations at least to some degree (e.g. nearly every expectation has at least two items) and is used as the acceptable level on this criterion. Index values between 0.6 and 0.7 indicate the Balance-of-Representation criterion has only been “weakly” met.

Source-of-Challenge Criterion

The Source-of-Challenge criterion is used to identify items for which the major cognitive demand is inadvertently placed and is other than the targeted reporting category or expectation (i.e. construct irrelevance). Bias and sensitivity issues, as well as technical issues and error, could all be reasons for an item to have a Source-of-Challenge problem. Such item characteristics may result in some students not answering an assessment item or answering an assessment item incorrectly even though they possess the understanding and skills being assessed.

Cutoffs for Alignment Criteria: Independent Analysis

For overall alignment, an assessment form is reported as *fully aligned* if no items need replacement to meet the conditions for all of the criteria described above. Note that “fully aligned” refers to the condition of meeting the *minimum acceptable levels* of alignment and does not mean that an assessment has “100% alignment” with the corresponding standards. A test form is considered *acceptably aligned* if it needs between one and five items replaced or revised in order to meet the minimum acceptable conditions for all alignment criteria. A test form is reported to *need slight adjustments* if six to ten items need to be replaced or revised to meet the minimum levels of alignment criteria and is reported to *need major adjustments* if more than ten items need to be replaced or revised. These categories represent typically used cutoff levels in the context of submission to federal peer review.

Confirmatory Analysis Criterion

For the confirmatory analysis of each test form’s metadata, panelists considered the internally coded standard and responded to the question “*To what extent does the item assess the content (expectations) within the internally coded standard(s)?*” using the following categories:

EXACTLY Note that the item does NOT need to assess every aspect of a standard, but it needs to be a direct (“exact”) measurement of a central aspect of the standard. A correct response to the item allows for a direct inference about student knowledge/skills/abilities as expressed in the standard.

PARTIALLY The item somewhat targets the expectations within the standard and it can be considered a majority match. A correct response to the item allows for some inference about student knowledge/skills/abilities as expressed in the standard.

MINIMALLY The item only very minimally targets the expectations within the standard – and it can be considered only a minority match. A correct response to the item allows for very little or very indirect inference about student knowledge/skills/abilities as expressed in the standard.

NOT AT ALL The item does not assess the expectations within the standard. No inference can be made about student knowledge/skills/abilities as expressed in the standard based on a correct response.

Cutoffs for Alignment Criterion: Confirmatory Analysis

For the confirmatory alignment criterion, it was expected that 100% of items at least partially target the expectations within the corresponding standard. However, a 100% cutoff does not allow for any human error or for any variation in professional opinion or understanding of task purpose. Therefore, a 90% confidence interval was used as the cutoff. In other words, at least 90% of items on each test event must be considered by a majority of reviewers to exactly or partially target the internally assigned standard. Of these, it was expected that at least 60% were considered an exact match. This acceptable level was defined by what would be required to ensure that a student’s test score was based on data from at least some items that were considered to be an exact measure of the targeted standard. Proficient students are expected to respond successfully to around 50% of the items on a particular test event. The selection of each item depends on the examinee’s ability and the CAT aims to maximize the precision of the student ability estimates.

Findings: Overall Program Capacity to Generate Aligned Assessments

Assessment blueprint information, statistical analyses of aggregate data from all administered test events in West Virginia as reported in the assessment technical report, and narrative information about the CAT algorithm were reviewed to determine the potential for generation of aligned test events. Actual alignment, however, is dependent on the specific items selected, and the extent to which these items measure what they are intended to measure. The findings reported from test documentation pertain to the potential for generating aligned test events. The findings reported from the independent and confirmatory analyses of sample test events pertain to the actual alignment of the specific set of test events. Consideration of all findings in relation to one another allows for an inference about the overall program capacity to generate assessments aligned with corresponding grade level standards. Therefore, findings on the overall program capacity to generate assessments aligned with corresponding standards is reported below by alignment criterion.

Categorical Concurrence: *Both ELA and Mathematics assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for all reporting categories.*

Aggregate data from all administered test events in West Virginia show that blueprint specifications are met at the reporting category level on all administered test events (with the exception of a small number of atypical cases in which students took the same grade-level test twice). With the exception of the Writing and Language domain for Grades 3-8 ELA, blueprints specify a minimum of 10 or more items for each domain. The Writing and Language domain specifies a minimum of eight Language items and one polytomous writing prompt, weighted at 10 points, based on a three-part rubric that targets multiple standards. The specification of 10 or more items from each assessed content domain (Reporting Category) along with the minimum eight Language items and single weighted writing prompt, is in excess of the typically accepted six-item expectation to meet the alignment criterion for Categorical Concurrence. Therefore, the blueprints and aggregate data from all administered test events suggest the potential for all test events to fully meet the alignment criterion of Categorical Concurrence for all Reporting Categories. Panelists agreed with the vast majority of the internally coded standards, suggesting that the assessment program has the capacity to meet the alignment criterion for Categorical Concurrence for all reporting categories for both ELA and Mathematics across Grades 3-8. Additional supporting evidence comes from results of the independent analyses, which showed all reporting categories for all grades and all test forms met this criterion.

DOK Consistency: *ELA assessments for Grades 3-6 and Mathematics assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for all reporting categories. ELA assessments for Grades 7-8 need slight adjustments to meet this criterion for the Literary Reading and Informational Reading reporting categories.*

The item selection algorithm does take Depth-of-Knowledge into account for item selection. Test blueprints provide information about the intended distribution of DOK on any test event but does not provide specific information about the distribution of item DOK as relates to the expectations for complexity of engagement specified in the standards (as represented by DOK assignments to standards). WVDE may wish to revisit the intended distribution of DOK provided on test blueprints in relation to the distribution of DOK in the corresponding grade-level standards. The frequency of

different DOK levels expected by the grade-level standards sometimes differed greatly from the frequency of items by DOK level specified in the blueprint. For example, around 70% of Grade 7 ELA standards had a DOK 3 but only around 25% of items on a test form were specified to be DOK 3. While it may still be possible to generate a test form that appropriately samples the complexity of student knowledge even if these distributions differ, adjustments may be warranted.

Reviewers did not, overall, agree with the internally coded DOK levels. For example, reviewers coded only a very limited number of Mathematics items as DOK 3 across all grades whereas the assessment blueprint for each grade expects at least 15% of items to be DOK 3. This disagreement may represent a difference in interpretation of complexity as described by the DOK language system – as only three Mathematics content standards (all in Grade 7) were considered DOK 3. Despite the disagreement, the independent analysis of all but one Grade 3-8 Mathematics test forms and all Grade 3-6 ELA test forms analyzed fully met the criterion of DOK Consistency. The independent analysis of Mathematics test forms found that 19 of the 20 forms analyzed met the criterion of DOK Consistency for all reporting categories. Only one form had one reporting category that did not meet DOK consistency, but it can be considered a minor alignment gap as it could be resolved with the revision or replacement of just two items. All seven test forms reviewed for Grades 3-6 ELA met DOK Consistency for all reporting categories. This suggests that although the internal interpretation of DOK levels may have been different, the Mathematics Grades 3-8 and ELA Grades 3-6 test items were, overall, requiring student engagement at the appropriate levels of complexity.

The consistency in findings from the independent analysis, which included at least one test event sampled from below proficiency for each grade and subject area, suggests that the assessment program has the capacity to meet the alignment criterion of DOK Consistency for all reporting categories for ELA across Grades 3-6 and for Mathematics across Grades 3-8.

All test events analyzed for ELA Grades 7 and 8 had weak or unmet DOK consistency for both reading reporting categories, requiring the revision or replacement of around seven items, averaged across forms. Some revisions are needed to ensure that the reading reporting categories for ELA Grades 7 and 8 meet the criterion of DOK Consistency.

Range-of-Knowledge: *Mathematics assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for all reporting categories. ELA assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for Literary Reading and Informational Reading reporting categories. ELA assessments for Grades 3-8 have the capacity to fully meet or very nearly meet this alignment criterion for the Writing and Language reporting categories, with consideration for comparison against an appropriate set of standards.* Test blueprints attend to Range of Knowledge to some extent by specifying the minimum and maximum number of items for each subcategory (cluster) of standards for each Reporting Category (domain). Aggregate data from all administered test events in West Virginia show that blueprint specifications are met at the cluster level on all administered test events. The results of the independent analyses showed that all Mathematics test forms for all grades fully met the criterion of Range of Knowledge for all reporting categories. The independent analyses showed that all ELA test forms fully

met the criterion of Range of Knowledge for the Literary Reading and Informational Reading reporting categories across all grades. The consistency of independent analysis results across test forms gives confidence in the inference that Range could be met for the assessment program for all reporting categories for Mathematics Grades 3-8 and for both reading reporting categories for ELA Grades 3-8.

For ELA, the Writing and Language (WL) reporting category showed weak or unmet Range, using the full set of WL standards approved by WVDE. The typically accepted cutoff for this criterion is that at least 50% of standards within a reporting category are assessed by one or more items on a test form. For each grade, around five (32%-40%) of the 15-16 WL standards included in the study were assessed. However, the total count of WL standards includes three separate writing standards for different text types (argument, informative/explanatory, and narrative) even though students are not expected to write essays for all three text types on the summative assessment. Two other standards were included in this reporting category that are often omitted by states for the context of summative assessment alignment: Standard 35, that includes a focus on use of the Internet for purposes including interacting and collaborating with others, and Standard 29, that includes an expectation for writing “routinely” over the course of the year. Additionally, Standard 24, which specifies work “[w]ith some guidance and support from peers and adults” may not be targeted and reviewers may hesitate to use in the context of a summative assessment that is expected to be completed independently. It may therefore be appropriate to omit these standards for alignment analysis in the context of a one-time, on-demand assessment that is expected to be completed independently. It may also be appropriate to consider the three Text Types and Purposes writing standards as a single mega-standard as the state does not expect students to complete more than one essay on the assessment.

If the Writing and Language reporting category were viewed as described above, it would include a total of 10-11 standards, resulting in all test events meeting or weakly meeting the Range-of-Knowledge criterion. However, even if the full set of standards is considered, including standards that are not intended be targeted on the summative assessment, unmet Range was still not considered a significant alignment issue for any test event analyzed, as it could be resolved with the revision or addition of just two or three items, and therefore does not affect the overall determination of acceptable alignment for test forms that fully meet the other alignment criteria.

Balance of Representation: *Mathematics assessments for Grades 3-8 have the capacity to fully meet the Balance criterion for all reporting categories. ELA assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for Literary Reading and Informational Reading reporting categories. Any weak Balance for ELA assessments for Grades 3-8 can be resolved along with Range, as described above and is not considered an alignment issue on its own.* The CAT algorithm attends to Balance of Representation with constraints to limit overemphasis of a particular standard. No overemphasis was evident in the sample test forms analyzed. All test forms for all grades in Mathematics and ELA met the Balance of Representation criterion for all reporting categories, with the exception of the Writing and Language reporting category. The weak balance shown there is connected to the Range of Knowledge issue described above and is not considered an alignment issue.

In summary, a consideration of all findings in relation to one another suggests that the West Virginia General Summative Assessment program has the capacity to generate assessments aligned with the corresponding grade level standards for all Grades 3-8 for Mathematics and for Grades 3-6 for ELA.

ELA Grades 7 and 8 standards and assessments met acceptable levels of alignment for three of the four alignment criteria. DOK Consistency was the only significant alignment gap. To meet DOK Consistency, from five to eight items would need to be revised or replaced (per test event). The assessment program could be expected to generate aligned test events for these grades with appropriate revisions in place.

Findings: ELA Grades 3-8 Item-Level Content Alignment Analysis of Sample Test Forms

Findings are reported for the 17 ELA test forms analyzed. For each grade, this included at least one test event sampled from below proficiency and at least one test event sampled from at proficiency. The elementary panel was able to analyze one test event sample from above proficiency (for Grade 3) and the middle grades panel was able to analyze a test sampled from above proficiency for Grade 6. All 11 ELA test forms reviewed for Grades 3-6 were found to be acceptably aligned with the corresponding grade level standards. The only criterion for alignment that was not fully met for the Grades 3-6 test forms—according to the typically used cutoff—was Range of Knowledge for the Writing and Language reporting category. For all test forms reviewed, weak or unmet Range could be resolved with the revision or addition of two or three 1-point items that targeted at least three additional standards within the Writing and Language reporting category that were otherwise unassessed. However, weak or unmet Range of Knowledge—according to the typically used cutoff for Range—could also be considered an artifact of the standards included rather than an alignment issue. For example, while the writing standards include separate standards for argumentative, informative/explanatory, and narrative writing, per test design and assessment blueprint, WVDE does not expect that students complete three separate essays or target all three types of writing in a single response. Additionally, three other standards included in the study (which reference guidance from adults, online collaboration, and routine writing over time) are not usually appropriate assessment targets for an on-demand assessment completed individually. If the test forms were considered against a reduced set of Writing and Language standards (with these standards combined/excluded), they would all meet or weakly meet the criterion of Range of Knowledge. Therefore, any unmet Range for Writing and Language domain (when compared against the full set of Writing and Language standards) is not considered an alignment issue.

All writing prompts were considered to address multiple standards within the Writing and Language reporting category. Based on the three-part rubric, student responses are evaluated in relation to both Writing standards and Language standards within the reporting category. The writing prompt is weighted at 10 points total and scored with a three-part rubric: Purpose, Focus, and Elaboration (4 points); Evidence and Elaboration (4 points), and Conventions of Standard English (2 points).

For all sample test events analyzed for ELA Grades 7 and 8, each reporting category met most alignment criteria but forms were found to need slight adjustments to meet the minimum cutoffs for acceptable alignment. The main alignment issue for Grades 7 and 8 ELA test events was unmet or weak DOK Consistency for both Literary Reading and Informational Reading reporting categories on all ELA test events analyzed for these grades. Averaged across forms, addressing unmet DOK consistency would require the revision or replacement of about seven items per test form.

Standards

The West Virginia ELA content standards were reorganized slightly for the test context. This rearranged structure reflected the structure of the standards as used on the West Virginia Department of Education's assessment blueprints. The West Virginia reading standards were organized by strand (Key Ideas and Details, Craft and Structure, etc.) but were separated out into a reporting category for Literary Text and another reporting category for Informational Text in the assessment blueprints. The blueprint organization of standards also combined the Writing and Language domains into a single reporting category.

A summary of the levels of complexity within the West Virginia College- and Career-Readiness Standards for Grades 3-8 ELA is given in **Table 1**, on the following page. Only one or two of the standards included in the study for each grade (3% or 6%) were considered DOK 1. These expectations targeted command of conventions of Standard English. Between eight and 17 standards (24% - 52%) per grade were considered a DOK 2, emphasizing work that involves both comprehension and subsequent processing of text, making basic inferences from text and using specific information from text to explain events and ideas, as well as purposeful application of language knowledge and skills. Around half of the Grades 3-6 standards were considered DOK 2. A smaller proportion (24% or 27%) of the Grades 7 and 8 standards were considered DOK 2. Panelists thought that the language of many of the reading standards shifted between Grades 6 and 7 to higher complexity expectations (DOK 3 vs 2). For example, in Grade 6, Standard 3 emphasized that students describe a sequence of events as conveyed within the text while in Grade 7, the expectation in Standard 3 goes beyond what is represented in the text and asks students to analyze how different aspects of the text influence one another. Around 40% of the Grades 3-6 standards and around 70% of the Grades 7 and 8 standards were considered to be DOK 3. These DOK 3 standards emphasized expectations for deep analysis of text and abstract thinking, including making holistic inferences based on text, and engaging in critical reading to consider aspects of author's purpose and use of textual features. These DOK 3 standards also included expectations for argumentative, informative/explanatory, and narrative writing as well as other expectations related to the process of writing. For each grade, only standard 29 was considered DOK 4. A DOK 4 expectation is one that is at least as complex as a DOK 3 but also requires extended time—days, weeks, or months—to complete. Standard 29 includes an expectation for writing that involves research, reflection, and revision over extended time frames. Although some components of these DOK 4 standards may be reasonably assessed by on-demand assessments, DOK 4 standards should not be expected to be fully assessed by an on-demand assessment. All of the expectations used in this study will be referred to as *standards* in this report.

It may be appropriate to omit Standards 24, 25, and 29 when evaluating the alignment of the independently completed summative assessment, because these standards emphasize work such as collaboration, support from adults, and ongoing routine writing, which is not targeted in this context. Additionally, the argumentative, informative/explanatory, and narrative writing standards might be best considered as a single aggregate writing standard, as students are expected per assessment blueprint to write just one essay in the context of the assessment. Results are shown as relates to both the extended and reduced set of Writing and Language standards in this section.

Table 1. Expectations by Depth-of-Knowledge (DOK) Levels for West Virginia ELA Standards used in the Grades 3-8 ELA Alignment Analysis, 2019

ELA	Total Number of Expectations	DOK Level	Number of Standards by Level	Percent within Grade by Level
Grade 3	32	1	2	6
		2	15	47
		3	14	44
		4	1	3
Grade 4	33	1	1	3
		2	17	52
		3	14	42
		4	1	3
Grade 5	33	1	1	3
		2	16	48
		3	15	45
		4	1	3
Grade 6	33	1	2	6
		2	17	52
		3	13	39
		4	1	3
Grade 7	33	1	1	3
		2	8	24
		3	23	70
		4	1	3
Grade 8	33	1	1	3
		2	9	27
		3	22	67
		4	1	3

Test Forms

The West Virginia ELA assessment for Grades 3-8 consisted of 35-41 machine scorable items per test event and a writing prompt. The machine-scored items were one or two points each. The writing prompt was weighted at 10 points total and scored with a three-part rubric with descriptions of expectations for the following categories: Purpose, Focus, and Elaboration (4 points); Evidence and Elaboration (4 points), and Conventions of Standard English (2 points). No field test items were included in the analysis. The West Virginia ELA assessments included 10 different item types, named and described in **Table 2**, on the following page.

Table 2. Item Types for West Virginia Computer Adaptive General ELA Assessments for Grades 3-8

Response Type	Description
Evidence-Based Selected Response (EBSR)	Student selects the correct answers from Part A and Part B. Part A often asks the student to make an analysis or inference, and Part B requires the student to use text to support Part A.
Extended Response (ER)	Student is directed to provide a longer, written response.
Editing Task Choice (ETC)	Student identifies an incorrect word or phrase and chooses the replacement from a number of options.
Grid (GI)	Student selects words, phrases, or images and uses the drag-and-drop feature to place them into a graphic organizer.
Hot Text (HT)	Student is directed to either select or use the drag-and-drop feature to use text to support an analysis or make an inference.
Matching (MI)	Student checks a box to indicate if information from a column header matches information from a row.
Multiple Choice (MC)	Student selects one correct answer from a number of options.
Multiple Choice/Select + Hot Text (Two-Part HT)	Student selects the correct answer from Part A and Part B. Part A is multiple choice or multiple select and Part B is hot text.
Multiple Select (MS)	Student selects all correct answers from a number of options.
Natural Language (NL)	Student uses the keyboard to enter a response into a text field.

Source: WVGS A 2018-2019 Technical Report: Volume 1, p. 11

If no particular grade-level standard was targeted by a given assessment item, reviewers were instructed to code the item at a “higher” or more inclusive level, such as the strand level for ELA or domain level for Mathematics. This coding to a “generic standard” sometimes indicates that the item was inappropriate for a particular grade level (for example, the item might better match a standard from another grade level). If the item was grade-appropriate but an appropriate standard was not found, a generic coding may indicate that there is a part of the content within the standards that is being interpreted differently by different parties. Reviewers found that nearly all of the test items reasonably addressed specific standards. No generic standards were used for 14 of the 17 ELA test forms reviewed. On just three of the 17 ELA test forms analyzed, a majority of reviewers mapped only one assessment item per form to a generic standard, indicating that the item did not directly target the content within any of the standards.

Table 3, on the next page, shows the items for each test form that a majority of reviewers coded to a generic standard. This table shows the generic standard to which the item was coded, the number of reviewers who coded the item to the generic standard, and the reason for the coding. Reviewers were required to write an explanation in the case of assigning an item to a generic standard. These notes can be found in **Appendix D** (provided to the state and to the assessment vendor but redacted for public release). Items assigned to generic standards by a majority of panelists should be reviewed. It is possible that these items are inappropriately placed or coded in the item bank.

Table 3. Items Assigned to Generic Content Expectations by Assessment by a Majority of Reviewers, West Virginia ELA Alignment Analysis, 2019

ELA Grade/Form	Generic Content Expectation	Item Number(# of Reviewers)	Reason
Grade 6 Batch 1	LR.6	16(5)	Panelists commented that this item relates in some way to expectations within standards 1, 9, 14, and 40 but that none of these standards was directly assessed.
Grade 6 Batch 2	IR.6.IK1	30(4)	Panelists noted that the item required consideration across passages but not across different types of media; that the answer choices were statements of facts and not claims, and that they eliminated the options, finding no standard that the item directly addressed.
Grade 8 Batch 1	IR.8.CS	24(4)	Panelists commented that the item was related to craft and structure but that they did not find a specific standard that the item directly targeted.

Alignment Statistics and Findings for ELA Grades 3-8

Overall alignment results are summarized in **Table 4** on the next page and then detailed for each test form in the pages that follow. Based on typically accepted cutoffs for the four main alignment criteria considered in this study, all test forms analyzed for Grades 3-6 would be considered acceptably aligned with West Virginia College- and Career-Readiness Standards for Grades 3-6 ELA. For all test events analyzed for ELA Grades 7 and 8, all reporting categories met most alignment criteria but five out of the six test events analyzed were found to need slight adjustments to meet the minimum cutoffs for overall alignment. To be considered fully aligned, each test event required the revision or replacement of around 7 items per test form to address weak or unmet DOK consistency for both reading reporting categories.

Across Grades 3-8, if considering the reduced set of standards commonly included in the Writing and Language reporting category (omitting standards that expect collaborative work and combining the Text Types and Purposes cluster of writing standards), four test events would meet Range for the WL Reporting Category and the other test events would weakly meet Range for WL, requiring only one or two items replaced or added. Even if considering the extended set of standards, only two or three items would need to be replaced or added for each test form to meet Range for the WL Reporting Category.

All test events analyzed met the expectation that an independent analysis by an expert educator panel agreed that at least 90% of items exactly or partially targeted a central aspect of the corresponding internally assigned standards (i.e. the intended assessment targets).

Overall summary findings are shown in **Table 4**, which continues on the next page. Results for the independent analysis are included for both the extended set of WL standards and for the reduced set of WL standards.

Table 4. Overall Alignment Findings for Grades 3-8 ELA Assessments with West Virginia College- and Career-Readiness Standards for Grades 3-8 ELA

Test Form	Independent Analysis Findings		Approx. # of Items that Need Revision/Replacement for Full Alignment		Confirmatory Analysis Findings
	Extended Set of WL Standards	Reduced Set of WL Standards	Extended Set of WL Standards	Reduced Set of WL Standards	
Grade 3 B1	Acceptably Aligned	Fully aligned	3 items	--	Fully meets
Grade 3 B2	Acceptably Aligned	Fully aligned	3 items	--	Fully meets
Grade 3 B4	Acceptably Aligned	Acceptably Aligned	4 items	1 item	NT*
Grade 4 B1	Acceptably Aligned	Fully aligned	2 items	--	NT
Grade 4 B2	Acceptably Aligned	Fully aligned	2 items	--	Fully meets
Grade 5 B1	Acceptably Aligned	Acceptably Aligned	3 items	1 item	NT
Grade 5 B2	Acceptably Aligned	Acceptably Aligned	3 items	1 item	Fully meets
Grade 6 B1	Acceptably Aligned	Acceptably Aligned	3 items	1 item	Fully meets
Grade 6 B2 - ELEM PANEL	Acceptably Aligned	Acceptably Aligned	3 items	1 item	Fully meets

*Not Tested

Table 4 Cont'd. Overall Alignment Findings for Grades 3-8 ELA Assessments with West Virginia College- and Career-Readiness Standards for Grades 3-8 ELA

Test Form	Independent Analysis Findings		Approx. # of Items that Need Revision/Replacement for Full Alignment		Confirmatory Analysis Findings
	Extended Set of WL Standards	Reduced Set of WL Standards	Extended Set of WL Standards	Reduced Set of WL Standards	
Grade 6 B2 - MS PANEL	Acceptably Aligned	Fully aligned	2 items	--	Fully meets
Grade 6 B3	Acceptably Aligned	Acceptably Aligned	3 items	1 item	Fully meets
Grade 6 B4	Acceptably Aligned	Fully aligned	2 items**	--	NT*
Grade 7 B1	Needs Slight Adjustments	Needs Slight Adjustments	9 items	7 items	Fully meets
Grade 7 B2	Needs Slight Adjustments	Acceptably Aligned	7 items	5 items	Fully meets
Grade 7 B3	Needs Major Adjustments	Needs Slight Adjustments	11 items**	8 items	Fully meets**
Grade 8 B1	Needs Slight Adjustments	Needs Slight Adjustments	9 items	7 items	Fully meets
Grade 8 B2	Needs Slight Adjustments	Needs Slight Adjustments	10 items	7 items	Fully meets
Grade 8 B3	Needs Slight Adjustments	Needs Slight Adjustments	10 items	7 items	Fully meets**

*Not Tested

**Coded by three reviewers

Two or more test events were analyzed per grade. Some items were used in more than one test event (repeated across test events). In these cases, reviewers' previous codings repopulated for efficiency in the item viewing interface (i.e., their "test booklet"). For each grade, between 67 and 109 unique items were reviewed in aggregate. This represented between 17% and 34% of the item pool for that grade. Overall, a total of 22% of the Grades 3-8 ELA item pool was analyzed in this study. The number and proportion of items reviewed per grade is detailed in **Table 5**, on the following page. There was very little variation of results across ELA test forms analyzed, suggesting that additional test events would likely yield similar results. All ELA test forms reviewed for Grades 3-6 showed similar results and all ELA test forms reviewed for Grades 7-8 showed similar results.

Table 5. Number of Unique Items Analyzed and Proportion of Grade Level Item Bank by Grade – West Virginia Grades 3-8 ELA Assessments

Grade	Number of unique items analyzed	Total operational item count	Proportion of item bank analyzed
3	109	325	34%
4	67	397	17%
5	67	361	18%
6	104	455	23%
7	104	472	22%
8	94	439	21%
TOTAL (ELA GRADES 3-8)	545	2449	22%

Results by Test Form – Independent Analysis

The results of the independent item-level analysis for each of the four alignment criteria are provided in **Tables 6 - 23** for each ELA test event analyzed for the reporting categories of Reading Literary Text (RL), Reading Informational Text (RI), and Writing and Language (WL). The approximate numbers of replaced or revised items necessary to meet minimum levels of alignment are provided for each test form. More detailed data on each of the criteria are given in **Appendix B**, in the first three tables for each test form.

In **Tables 6 - 23**, “YES,” indicates that an acceptable level was attained between the assessment and the reporting category on the criterion. “WEAK” indicates that the criterion was nearly met, within a margin that could simply be due to error or reasonable variation in reviewer coding. “NO” indicates that the criterion was not met by a noticeable margin—10% under an acceptable level for Depth-of-Knowledge Consistency, 10% under an acceptable level for Range-of-Knowledge Correspondence, and 0.1 under an index value of 0.7 for Balance-of-Representation. Categorical Concurrence is reported in average number of items or assigned points. Depth-of-Knowledge Consistency is reported by the percent of items that were at or above the DOK of the corresponding standard. Range-of-Knowledge is reported as the percent of standards within each reporting category that were targeted by one or more items. Balance-of-Representation is an index value, ranging from 0-1.

In **Tables 6 - 23**, results for Range for WL are given in relation to the reduced set of WL standards. The corresponding results if using the expanded set of WL standards are shown in parentheses.

Panelists reviewed a minimum of two test forms per grade. As time allowed, they reviewed additional forms.

ELA Grade 3: Test Forms G3B1, G3B2, and G3B4

Panelists reviewed three test forms, one sampled from below proficiency (Form G3B1), one sampled from at proficiency (Form G3B2), and one sampled from above proficiency (G3B4). All three test forms showed similar results and are considered fully aligned or acceptably aligned with the Grade 3 ELA standards (no matter if considering the extended or condensed set of Writing and Language standards). All test forms met all alignment criteria for the Reading Literary Text reporting category and for the Reading Informational Text reporting category. All forms met or weakly met Range-of-Knowledge for the Writing and Language reporting category if considering the condensed set of standards. All forms had unmet Range-of-Knowledge for the Writing and Language reporting category if considering the extended set of standards. If using the extended set of standards, test forms would need three (Forms G3B1 and G3B2) or four (Form G3B4) one-point items revised, replaced, or added to meet the minimum levels of acceptable alignment. The revisions would need to target at least three or four additional standards within the Writing and Language reporting category. If considering the condensed set of Writing and Language standards, Forms G3B1 and G3B2 would be considered fully aligned and just one item would need to be revised, replaced, or added on Form G3B4 added to be considered fully aligned with the Grade 3 ELA standards. The revisions would need to target at least one additional standard within the Writing and Language reporting category.

Table 6. Results for ELA Grade 3 Form B1 and West Virginia College- and Career-Readiness Standards for ELA Grade 3

WV Grade 3 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.3 Grade 3 Reading Literary Text	23.3	67%	81%	0.76	YES	YES	YES	YES
RI.3 Grade 3 Reading Informational Text	9.2	65%	66%	0.84	YES	YES	YES	YES
W.L.3 Grade 3 Writing and Language	25.7	100%	51% (34%)	0.69	YES	YES	YES (NO)	WEAK

*Number of items/points
(if using extended set of WL standards)

Table 7. Results for ELA Grade 3 Form B2 and West Virginia College- and Career-Readiness Standards for ELA Grade 3

WV Grade 3 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.3 Grade 3 Reading Literary Text	17.8	72%	95%	0.81	YES	YES	YES	YES
RI.3 Grade 3 Reading Informational Text	14.0	70%	85%	0.75	YES	YES	YES	YES
W.L.3 Grade 3 Writing and Language	28.2	100%	51% (34%)	0.7	YES	YES	YES (NO)	YES

*Number of items/points
(if using extended set of WL standards)

Table 8. Results for ELA Grade 3 Form B4 and West Virginia College- and Career-Readiness Standards for ELA Grade 3

WV Grade 3 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.3 Grade 3 Reading Literary Text	18.0	73%	85%	0.71	YES	YES	YES	YES
RI.3 Grade 3 Reading Informational Text	14.5	77%	68%	0.83	YES	YES	YES	YES
W.L.3 Grade 3 Writing and Language	26.0	86%	48% (32%)	0.74	YES	YES	WEAK (NO)	YES

*Number of items/points
(if using extended set of WL standards)

ELA Grade 4: Test Forms G4B1 and G4B2

Panelists reviewed two Grade 4 ELA test forms, one sampled from below proficiency (Form G4B1) and one sampled from at proficiency (Form G4B2). Both test forms showed similar results and are considered fully aligned or acceptably aligned with the Grade 4 standards (no matter if considering the extended or condensed set of Writing and Language standards). Both test forms met all alignment criteria for the Reading Literary Text reporting category and all criteria for the Reading Informational Text reporting category. Both forms met Range-of-Knowledge for the Writing and Language reporting category if considering the condensed set of standards but had weak or unmet Range-of-Knowledge for the Writing and Language reporting category if considering the extended set of standards. If using the extended set of standards, for both test forms, two one-point items would need to be revised, replaced, or added to meet the minimum levels of acceptable alignment. The revisions would need to target at least two additional standards within the Writing and Language reporting category.

Table 9. Results for ELA Grade 4 Form B1 and West Virginia College- and Career-Readiness Standards for ELA Grade 4

WV Grade 4 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.4 Grade 4 Reading Literary Text	17.5	83%	95%	0.83	YES	YES	YES	YES
RI.4 Grade 4 Reading Informational Text	13.5	72%	68%	0.79	YES	YES	YES	YES
W.L.4 Grade 4 Writing and Language	28.8	76%	54% (40%)	0.72	YES	YES	YES (WEA K)	YES

*Number of items/points
(if using extended set of WL standards)

Table 10. Results for ELA Grade 4 Form B2 and West Virginia College- and Career-Readiness Standards for ELA Grade 4

WV Grade 4 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.4 Grade 4 Reading Literary Text	16.2	84%	83%	0.86	YES	YES	YES	YES
RI.4 Grade 4 Reading Informational Text	14.3	80%	79%	0.78	YES	YES	YES	YES
W.L.4 Grade 4 Writing and Language	28.5	70%	51% (38%)	0.66	YES	YES	YES (NO)	WEAK

*Number of items/points
(if using extended set of WL standards)

ELA Grade 5: Test Forms G5B1 and G5B2

Panelists reviewed two test forms, one sampled from below proficiency (Form G5B1) and one sampled from at proficiency (Form G5B2). Both test forms showed similar results and are considered acceptably aligned with the Grade 5 standards (no matter if considering the extended or condensed set of Writing and Language standards). Both test forms met all alignment criteria for the Reading Literary Text reporting category and all criteria for the Reading Informational Text reporting category. Both forms weakly met Range-of-Knowledge for the Writing and Language reporting category if considering the condensed set of standards but had unmet Range-of-Knowledge for the Writing and Language reporting category if considering the extended set of standards.

If considering the condensed set of standards, just one item would need to be revised, replaced, or added for each test form to be considered fully aligned. If using the extended set of standards, for both test forms, three one-point items would need to be revised, replaced, or added to meet the minimum levels of acceptable alignment. The revisions would need to target at least three additional standards within the Writing and Language reporting category.

Table 11. Results for ELA Grade 5 Form B1 and West Virginia College- and Career-Readiness Standards for ELA Grade 5

WV Grade 5 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.5 Grade 5 Reading Literary Text	17.3	84%	77%	0.81	YES	YES	YES	YES
RI.5 Grade 5 Reading Informational Text	15.2	92%	87%	0.79	YES	YES	YES	YES
W.L.5 Grade 5 Writing and Language	27.3	81%	47% (35%)	0.76	YES	YES	WEAK (NO)	YES

*Number of items/points
(if using extended set of WL standards)

Table 12. Results for ELA Grade 5 Form B2 and West Virginia College- and Career-Readiness Standards for ELA Grade 5

WV Grade 5 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.5 Grade 5 Reading Literary Text	17.3	86%	79%	0.80	YES	YES	YES	YES
RI.5 Grade 5 Reading Informational Text	13.3	91%	74%	0.85	YES	YES	YES	YES
W.L.5 Grade 5 Writing and Language	27.2	77%	44% (33%)	0.69	YES	YES	WEAK (NO)	WEAK

*Number of items/points
(if using extended set of WL standards)

ELA Grade 6: Test Forms G6B1, G6B2, G6B3, and G6B4

Panelists reviewed four Grade 6 test forms. Both the elementary and middle school panels analyzed Form G6B2, to check on consistency between panels in their coding. The two panels' codings yielded similar alignment findings. The middle school panel additionally analyzed Form G6B1 (below proficiency), Form G6B3 (at proficiency), and Form G6B4 (above proficiency). Form G6B4 was reviewed by three panelists. All test forms showed similar results and are considered fully aligned or acceptably aligned with the Grade 6 standards (no matter if considering the extended or condensed set of Writing and Language standards). All test forms met all alignment criteria for the Reading Literary Text reporting category and all criteria for the Reading Informational Text reporting category. All forms met or weakly met Range-of-Knowledge for the Writing and Language reporting category if considering the condensed set of standards but had unmet Range-of-Knowledge for the Writing and Language reporting category if considering the extended set of standards.

If using the condensed set of standards, all forms are fully or acceptably aligned and any alignment weakness could be resolved with the revision, replacement, or addition of just one item that addressed one additional Writing and Language standard. If using the extended set of standards, for all test forms, three one-point items would need to be revised, replaced, or added to meet the minimum levels of acceptable alignment. The revisions would need to target at least three additional Writing and Language standards. Revisions to address Range also would likely resolve any weakness in Balance. If other alignment criteria are met, weak Balance is typically not considered an issue.

Table 13. Results for ELA Grade 6 Form B1 and West Virginia College- and Career-Readiness Standards for ELA Grade 6

WV Grade 6 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.6 Grade 6 Reading Literary Text	12.6	100%	71%	0.75	YES	YES	YES	YES
RI.6 Grade 6 Reading Informational Text	17.0	79%	86%	0.77	YES	YES	YES	YES
W.L.6 Grade 6 Writing and Language	25.4	100%	43% (32%)	0.66	YES	YES	WEAK (NO)	WEAK

*Number of items/points
(if using extended set of WL standards)

Table 14. Results for ELA Grade 6 Form B2 and West Virginia College- and Career-Readiness Standards for ELA Grade 6 – Elementary Panel

WV Grade 6 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.6 Grade 6 Reading Literary Text	12.6	90%	75%	0.88	YES	YES	YES	YES
RI.6 Grade 6 Reading Informational Text	17.0	73%	77%	0.8	YES	YES	YES	YES
W.L.6 Grade 6 Writing and Language	25.4	77%	45% (33%)	0.77	YES	YES	WEAK (NO)	YES

*Number of items/points
(if using extended set of WL standards)

Table 15. Results for ELA Grade 6 Form B2 and West Virginia College- and Career-Readiness Standards for ELA Grade 6 – Middle School Panel

WV Grade 6 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.6 Grade 6 Reading Literary Text	13.2	94%	87%	0.84	YES	YES	YES	YES
RI.6 Grade 6 Reading Informational Text	18.4	78%	75%	0.76	YES	YES	YES	YES
W.L.6 Grade 6 Writing and Language	27.2	98%	50% (37%)	0.73	YES	YES	YES (NO)	YES

*Number of items/points
(if using extended set of WL standards)

Table 16. Results for ELA Grade 6 Form B3 and West Virginia College- and Career-Readiness Standards for ELA Grade 6

WV Grade 6 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.6 Grade 6 Reading Literary Text	15.0	91%	75%	0.81	YES	YES	YES	YES
RI.6 Grade 6 Reading Informational Text	18.6	93%	82%	0.82	YES	YES	YES	YES
W.L.6 Grade 6 Writing and Language	27.0	99%	46% (35%)	0.69	YES	YES	YES (NO)	WEAK

*Number of items/points
(if using extended set of WL standards)

Table 17. Results for ELA Grade 6 Form B4 and West Virginia College- and Career-Readiness Standards for ELA Grade 6 (Three Reviewers)

WV Grade 6 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.6 Grade 6 Reading Literary Text	14.3	90%	75%	0.79	YES	YES	YES	YES
RI.6 Grade 6 Reading Informational Text	18.7	91%	81%	0.79	YES	YES	YES	YES
W.L.6 Grade 6 Writing and Language	27.7	76%	50% (37%)	0.69	YES	YES	YES (NO)	WEAK

*Number of items/points
(if using extended set of WL standards)

ELA Grade 7: Test Forms G7B1, G7B2, and G7B3

All panelists reviewed two test forms, one sampled from below proficiency (Form G7B1) and one sampled from at proficiency (Form G7B2). An analysis of Form G7B3 (at proficiency) was completed by three reviewers. All test forms showed similar results. If considering the condensed set of standards, Form G7B2 would be considered acceptably aligned and Forms G7B1 and G7B3 would be considered to need slight adjustments. If considering the expanded set of standards, Form G7B1 and G7B2 would be considered to need slight adjustments and Form G7B3 would need major adjustments.

If considering the condensed set of standards, three of the four alignment criteria were met or weakly met for both Reading reporting categories and all alignment criteria were met or weakly met for the Writing and Language reporting category. The alignment gap for all test forms was weak or unmet DOK Consistency for the Reading Literary (RL) Text and Reading Informational (RI) Text reporting categories. To meet DOK Consistency for the RL reporting category, two or three items/points would need to be revised, replaced, or added. Similarly for the RI reporting category, five items/points (Forms G7B1 and G7B3), or two items/points (Form G7B2) would need to be revised, replaced, or added to meet DOK Consistency. These items would all need to have a DOK that matched the DOK of the targeted standards.

If using the extended set of standards, two items/points (Forms G7B1 and G7B2) or three items/points (Form G7B3) would need to be revised, replaced, or added that targeted at least three additional standards within the Writing and Language reporting category to meet Range. If using the condensed set of standards, Range was met or weakly met for all forms. Revisions to address Range would also likely resolve any weakness in Balance. If other alignment criteria are met, weak Balance is typically not considered an issue.

If using the condensed set of standards, any alignment weakness could be resolved with the revision, replacement, or addition of five to eight one-point items. If using the extended set of standards, seven to 11 one-point items would need to be revised, replaced, or added to meet the minimum levels of acceptable alignment.

Table 18. Results for ELA Grade 7 Form B1 and West Virginia College- and Career-Readiness Standards for ELA Grade 7

WV Grade 7 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.7 Grade 7 Reading Literary Text	14.0	37%	82%	0.81	YES	NO	YES	YES
RI.7 Grade 7 Reading Informational Text	17.4	25%	86%	0.77	YES	NO	YES	YES
W.L.7 Grade 7 Writing and Language	28.6	84%	50% (37%)	0.74	YES	YES	YES (NO)	YES

*Number of items/points
(if using extended set of WL standards)

Table 19. Results for ELA Grade 7 Form B2 and West Virginia College- and Career-Readiness Standards for ELA Grade 7

WV Grade 7 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.7 Grade 7 Reading Literary Text	13.4	32%	70%	0.79	YES	NO	YES	YES
RI.7 Grade 7 Reading Informational Text	17.8	41%	73%	0.78	YES	WEAK	YES	YES
W.L.7 Grade 7 Writing and Language	27.0	83%	50% (37%)	0.66	YES	YES	YES (NO)	WEAK

*Number of items/points
(if using extended set of WL standards)

Table 20. Results for ELA Grade 7 Form B3 and West Virginia College- and Career-Readiness Standards for ELA Grade 7 (Three Reviewers)

WV Grade 7 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.7 Grade 7 Reading Literary Text	13.3	32%	83%	0.81	YES	NO	YES	YES
RI.7 Grade 7 Reading Informational Text	16.3	29%	81%	0.82	YES	NO	YES	YES
W.L.7 Grade 7 Writing and Language	25.3	74%	47% (35%)	0.74	YES	YES	WEAK (NO)	YES

*Number of items/points
(if using extended set of WL standards)

ELA Grade 8: Test Forms G8B1, G8B2, and G8B3

Panelists reviewed three test forms, one sampled from below proficiency (Form G8B1), and two sampled from at proficiency (Forms G8B2 and G8B3). All three test forms showed similar results. All forms would be considered to need slight adjustments for full alignment with the Grade 8 standards. If using the condensed standards, around seven items/points would need to be revised, added, or replaced per test form to meet the minimum levels of acceptable alignment according to the criteria used in this analysis. If using the expanded standards, nine or 10 items/points would need to be revised, added, or replaced per test form to meet the minimum levels of acceptable alignment.

If considering the condensed set of standards, three of the four alignment criteria were met or weakly met for both Reading reporting categories and all alignment criteria were met or weakly met for the Writing and Language reporting category. The alignment gap for all test forms was weak or unmet DOK Consistency for the Reading Literary (RL) Text and Reading Informational (RI) Text reporting categories. To meet DOK Consistency for the RL reporting category, three items/points would need to be revised, replaced, or added. Similarly for the RI reporting category, four items/points would need to be revised, replaced, or added to meet DOK Consistency. These items would all need to have a DOK that matched the DOK of the targeted standards.

If using the extended set of standards, two items/points (Form G8B1) or three items/points (Forms G8B2 and G8B3) would need to be revised, replaced, or added that targeted at least three additional standards within the Writing and Language reporting category to meet Range. If using the condensed set of standards, Range was met or weakly met for all forms.

Table 21. Results for ELA Grade 8 Form B1 and West Virginia College- and Career-Readiness Standards for ELA Grade 8

WV Grade 8 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.8 Grade 8 Reading Literary Text	14.2	28%	75%	0.71	YES	NO	YES	YES
RI.8 Grade 8 Reading Informational Text	16.0	27%	91%	0.81	YES	NO	YES	YES
W.L.8 Grade 8 Writing and Language	32.4	83%	50% (37%)	0.74	YES	YES	YES (NO)	YES

*Number of items/points
(if using extended set of WL standards)

Table 22. Results for ELA Grade 8 Form B2 and West Virginia College- and Career-Readiness Standards for ELA Grade 8

WV Grade 8 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.8 Grade 8 Reading Literary Text	12.4	30%	72%	0.76	YES	NO	YES	YES
RI.8 Grade 8 Reading Informational Text	18.6	30%	84%	0.8	YES	NO	YES	YES
W.L.8 Grade 8 Writing and Language	30.2	86%	45% (33%)	0.81	YES	YES	WEAK (NO)	YES

*Number of items/points
(if using extended set of WL standards)

Table 23. Results for ELA Grade 8 Form B3 and West Virginia College- and Career-Readiness Standards for ELA Grade 8

WV Grade 8 ELA Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RL.8 Grade 8 Reading Literary Text	14.8	38%	76%	0.82	YES	NO	YES	YES
RI.8 Grade 8 Reading Informational Text	16.0	30%	83%	0.76	YES	NO	YES	YES
W.L.8 Grade 8 Writing and Language	32.2	86%	48% (38%)	0.79	YES	YES	WEAK (NO)	YES

*Number of items/points
(if using extended set of WL standards)

Results by Test Form — Confirmatory Analysis

For the confirmatory analysis of each test form's metadata, panelists considered the internally coded standard and responded to the question "*To what extent does the item assess the content (expectations) within the internally coded standard(s)?*" using the following categories:

- **EXACTLY** (Note that the item does NOT need to assess every aspect of a standard, but it needs to be a direct measurement of a central aspect of the standard. A correct response to the item allows for a direct inference about student knowledge/skills/abilities as expressed in the standard.)
- **PARTIALLY** (The item somewhat targets the expectations within the standard and it can be considered a majority match. A correct response to the item allows for some inference about student knowledge/skills/abilities as expressed in the standard.)
- **MINIMALLY** (The item only very minimally targets the expectations within the standard – and it can be considered only a minority match. A correct response to the item allows for very little or very indirect inference about student knowledge/skills/abilities as expressed in the standard.)
- **NOT AT ALL** (The item does not assess the expectations within the standard. No inference can be made about student knowledge/skills/abilities as expressed in the standard based on a correct response.)

Panelists judged between 68% and 100% of machine-scorable items on each test form to be an exact measure of a central expectation within the internally coded standard. Panelists judged between 82% and 100% of machine-scorable items on all test forms to be an exact or partial measure of the internally coded standard. Results for all test forms are provided in **Table 24**, on the following page. In their notes, panelists provided their rationale(s) for any item that they thought only minimally addressed or did not at all address the internally coded standard. These notes can help to identify any items that are mislabeled or otherwise need adjustments or corrections.

Panelists noted that particularly for standards with significant overlap in expectations, they agreed that the item assessed the internally coded standard even if they had independently coded a different standard. This difference in coding reflects on the overlap in standards. For example, standards 7, 10, 39, and 40 all include an expectation for students to determine the meaning of unknown words in context. In concept, reviewers agreed that they could differentiate between and among the standards based on whether the word/phrase was presented in a Literary Reading context (Standard 7), an Informational Reading context (Standard 10) or a stand-alone type of context (Standard 39 or 40, depending on context). In practice, however, they found that it was not always such a clear distinction between the contexts. Therefore, even if a panelist had independently coded an item to Standard 7, for example, the panelist could also agree that the item reasonably targeted Standard 39. Panelists also reported some differences in perspective (between and among the assessment/AIR, WVDE, and panelists) about what constitutes literature vs informational text. Because the test blueprint separates standards into Literature and Informational Text domains, panelists sometimes found that items considered to be Informational Text by panelists and/or WVDE was coded to one of the Literature standards in the internal metadata (or vice versa). This suggests some clarification between and among stakeholders might be warranted but does not affect alignment, as Categorical Concurrence was met for all test events analyzed.

For the confirmatory analysis, panelists approached the coding of the writing prompt in disparate ways but had a common rationale. Panelists agreed that the internally coded standard was, indeed, targeted but they noted that the writing prompt targeted many additional standards, including Standards 23, 36, 37, and 38. Some panelists therefore recorded agreement with the internal coding and some recorded disagreement with the internal coding but all noted the same reason: they agreed that the internally coded standard was appropriate but also wanted to see additional standards included. Because panelists used different ways of recording this information on the confirmatory pass, the confirmatory analysis data are reported for the machine-scorable items only. The machine-scorable items each targeted a single standard and did not cause coding confusion.

The internally coded DOK shown in the metadata uses a slightly different interpretation than the definitions for Webb's Depth-of-Knowledge. For example, the internal metadata labels the writing prompt as a DOK 4. Per Webb's DOK definitions, a DOK 4 applies to a task that is both complex and requires extended time such that it cannot be completed in one sitting. This extended time allows for extended planning, reflection, synthesis of research to build an argument, thesis, or narrative, iterative revisions, and other types of engagement that is not likely or not possible in a single sitting. DOK 4 expectations are

not expected to be assessed on a single on-demand test. External reviewers coded fewer items as DOK 3 than are represented in the blueprints and internal metadata. Despite the differences, the independent analysis found that DOK Consistency was met for all reporting categories on all test forms for ELA Grades 3-6. DOK Consistency was not met for both reading reporting categories for Grades 7-8 forms. WVDE may wish to review the specifications for distribution of DOK (“DOK Ranges”) given in the test blueprints in relation to the distribution of DOK in the corresponding grade-level standards. The test blueprint specifies around 25% of the items on any ELA test form are expected to be DOK 3. In contrast, around 70% of Grades 7 and 8 standards were considered DOK 3.

Table 24. Results for West Virginia ELA Grades 3-8 Test Forms Confirmatory Analysis of Internally Coded Standard (Intended Assessment Target)

WV ELA Test Events	Exact	Exact or Partially	Minimally and/or Not at All	Alignment Finding
Grade 3 B1	90%	97%	3%	Fully meets
Grade 3 B2	92%	100%	0%	Fully meets
Grade 4 B2	100%	100%	0%	Fully meets
Grade 5 B2	100%	100%	0%	Fully meets
Grade 6 B1	71%	97%	3%	Fully meets
Grade 6 B2 - ELEM PANEL	87%	97%	3%	Fully meets
Grade 6 B2 - MS PANEL	68%	82%	18%	Fully meets
Grade 6 B3	87%	100%	0%	Fully meets
Grade 7 B1	71%	92%	8%	Fully meets
Grade 7 B2	77%	95%	5%	Fully meets
Grade 7 B3*	87%	100%	0%	Fully meets
Grade 8 B1	79%	100%	0%	Fully meets
Grade 8 B2	85%	100%	0%	Fully meets
Grade 8 B3*	80%	97%	3%	Fully meets

*Coded by three reviewers

Table 25 (on the following page) shows the specific items for which reviewers disagreed with vendor-designated assessment target by marking that the item only minimally targeted the internally coded standard or did not at all target the standard.

Table 25. Items Considered to Have a Miscoded Standard by a Split or Majority of Panelists - West Virginia ELA Grades 3-8 Test Forms Confirmatory Analysis

Test Form	Item Number	Split/Majority Coding	Rationale
Grade 3 B1	#35 – 10199- 6599	Minimally (3) Not at all (1)	Panelists noted that it was not necessary to refer to particular parts of a passage or build on earlier sections of the text, as specified in 3.8 (intended assessment target) and that 3.3 was a better fit as the item focus was a basic character question.
Grade 6 B1	#16 – 10199- 7194	Minimally (1) Not at all (2)	Panelists commented that this item was tangentially related to standards 6.1, 6.9, 6.14 (intended assessment target), and/or 6.40 but did not think that it directly assessed any standard.
Grade 6 B2 – ELEM PANEL	#5 – 10199- 4365	Minimally (6)	Panelists commented that because the answer choices were not directly from the text, students were not demonstrating work related to 6.11 (intended assessment target) which specifies analysis of a particular component of the text. Instead, panelists thought the item more directly assessed 6.12.
Grade 6 B2 – MIDDLE PANEL	#2 – 10199- 4362	Not at all (6)	Panelists did not agree that the item assessed standard 6.39 (intended assessment target) because although a Greek affix was referenced within the item, the commented that students did not need to use the Greek affix to answer the question.
	#5 – 10199- 4365	Minimally (4) Not at all (2)	Panelists commented that the item addressed purpose, which is part of standard 6.12, and not structure, which is the focus of 6.11 (intended assessment target).
	#7 – 10199- 4370	Minimally (6)	Panelists commented that the item focused more on claims and evidence than on analysis of a key person or event and that the student did not need to attend to how ideas are developed, as specified in 6.6 (intended assessment target).
	#16 – 10199- 6972	Minimally (6)	Panelists noted that the intended assessment target (6.40) specifies understanding or interpretation of figurative language but that the item only requires students to identify the figurative language.

Table 25 Cont'd. Items Considered to Have a Miscoded Standard by a Split or Majority of Panelists - West Virginia ELA Grades 3-8 Test Forms Confirmatory Analysis

Test Form	Item Number	Split/Majority Coding	Rationale
Grade 6 B2 – MIDDLE PANEL (Cont'd)	#27 – 10199- 7146	Minimally (5) Not at all (1)	Panelists commented that 6.17 (intended assessment target) expects students to compare and contrast two different presentations of events but that the item only requires student to select facts from a single passage.
	#30 – 10199- 7147	Minimally (6)	Panelists commented that 6.17 (intended assessment target) expects students to compare and contrast two different presentations of events but that the item asks students to combine information from passages, more like the expectations of 6.15.
Grade 7 B1	#4 – 10199- 8071	Minimally (3) Not at all (1)	Panelists commented that the question addressed only one passage while the intended assessment target expects analysis of two passages.
	#23 – REP1019 9-5116	Minimally (3)	Panelists commented that the item related to ideas in the targeted standard but that it did not require reading across multiple paragraphs to understand point of view of different characters as specified in 7.9.
	#36 – 10199- 7525	Minimally (2) Not at all (3)	This “mismatch” is an artifact of the study structure. The item targeted a Speaking and Listening standard that was not included in the study.
Grade 7 B2	#17 – 10199- 8061	Minimally (2) Not at all (1)	Panelists commented that the item requires students to make an inference or interpret a specific phrase.
	#24 – 10199- 5281	Minimally (4)	Panelists commented that the item focuses on three passages not just one.
Grade 8 B3	#11 – REP1019 9-8834	Minimally (2)*	Panelists commented that focus of the question related to foreshadowing, which wasn't directly related to the standards that addressed related topics such as structure, language, elements of a story or author's purpose.

*Review completed by three panelists

Source of Challenge Issues and Reviewers' Comments

Reviewers were instructed to document any Source-of-Challenge issue and to provide any other comments they may have about an item. A Source-of-Challenge is a technical issue with an item that can result in a student answering the item correctly or incorrectly for the wrong reason. Reviewers left comments in some Grades 6-8 Source-of-Challenge notes boxes, but none of the comments suggest the type of technical issue considered Source-of-Challenge. Nevertheless, all Source-of-Challenge comments should be reviewed as reviewers may have wanted to emphasize their comments by logging them in that space. Some comments logged in that space appear to simply be notes that were misplaced in the wrong text box. The full text of reviewers' notes and debriefing comments was provided to the state and vendor but has been redacted for public release.

Reviewers also wrote notes about a number of items. Some notes indicate when only part of a particular standard was targeted by an assessment task. These notes also include general comments about items. Some notes include suggestions for resolutions to issues identified. After coding each assessment form, reviewers were asked to respond to four debriefing questions. The main purpose of these debriefing questions was to provide a space for panelists to include any holistic comments or feedback that they did not have an opportunity to communicate in the item-level feedback. The full text of reviewers' notes and debriefing comments was provided to the state and vendor but has been redacted for public release.

Reliability among Reviewers

Reviewers engaged in some adjudication of their data after all reviewers finished their coding for an assessment. After the first two or three test events, broad differences in coding were limited and panels conducted very little or no adjudication. These discussions were used to identify any mistakes in coding. Reviewers were not required to change their coding after discussion unless they found a compelling reason. The agreement statistics shown in **Table 26**, on the following page, were computed after adjudication. An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the reviewers. The overall intraclass correlation among the ELA reviewers' assignment of DOK levels to items was high for all analyses, between 0.83 and 0.95 (**Table 26**).

A pairwise comparison was used to determine the degree of reliability of reviewers coding at the reporting category level and the standard level. The pairwise comparison was computed by considering for every item the coding assigned by each reviewer compared to the coding by each of the other five reviewers. For example, for six reviewers a total of 15 comparisons are computed for each item. For most alignment studies, the objective pairwise agreement is higher than 0.6. The pairwise agreement for assigning objectives to items met this benchmark for all studies and was very high for most studies in this analysis. For coding to the level of reporting category, a pairwise agreement of 0.90 is desired. For all but one test form, pairwise agreement for reporting category was above 0.90. Form G8B3 was close, with a pairwise agreement of 0.86. This statistic reflects differences in reviewer perspectives when a particular component of a learning expectation appeared in more than one reporting category. Reviewers commented on this issue in their notes, saying that there were a number of items for

which the group debated whether the item was a best fit for expectations within Reading or Language standards (as related to determining the meaning of a word in context). Reviewers discussed differentiating between the two expectations, came to some agreement on decision rules, but continued to have some differences in perspective as related to particular items.

Table 26. Intraclass and Pairwise Comparisons, West Virginia Grades 3-8 ELA Assessments with College- and Career-Readiness Standards for ELA Grades 3-8

Test Form	Intraclass Correlation (DOK)	Pairwise Comparison (DOK)	Pairwise Comparison (Reporting Category)	Pairwise Comparison (Objectives)
Grade 3 B1	0.94	0.70	0.98	0.75
Grade 3 B2	0.95	0.74	0.97	0.67
Grade 3 B4	0.95	0.72	0.97	0.75
Grade 4 B1	0.94	0.72	0.92	0.79
Grade 4 B2	0.94	0.74	0.93	0.70
Grade 5 B1	0.94	0.71	0.94	0.61
Grade 5 B2	0.93	0.70	0.97	0.63
Grade 6 B1	0.92	0.84	0.97	0.72
Grade 6 B2 - ELEM PANEL	0.84	0.72	0.94	0.65
Grade 6 B2 - MS PANEL	0.93	0.88	0.99	0.81
Grade 6 B3	0.91	0.86	0.94	0.69
Grade 6 B4	0.87	0.81	0.88	0.60
Grade 7 B1	0.90	0.82	0.92	0.68
Grade 7 B2	0.93	0.76	0.96	0.70
Grade 7 B3	0.90	0.77	0.97	0.74
Grade 8 B1	0.90	0.84	0.95	0.69
Grade 8 B2	0.90	0.77	0.90	0.63
Grade 8 B3	0.83	0.68	0.86	0.65

Findings: Mathematics Grades 3-8 Item-Level Content Alignment Analysis of Sample Test Forms

All 20 Mathematics test forms reviewed were found to be fully aligned (19 test forms) or acceptably aligned (one test form) with the corresponding grade level standards. One of the 20 test forms analyzed had two reporting categories with weak or unmet DOK Consistency. The test form was still considered acceptably aligned as the issues could be resolved with the revision or replacement of only two 1-point items that targeted the complexity of the corresponding standard.

Standards

The West Virginia College- and Career-Readiness Standards for Mathematics were reorganized slightly for the test context. This rearranged structure, which combines some domains for each grade, reflects the structure of the standards as used on the West Virginia Department of Education's assessment blueprints. A summary of the levels of complexity within the West Virginia College- and Career-Readiness Standards for Grades 3-8 Mathematics is given in **Tables 27 and 28**, on the following page. Between 8% (Grade 7) and 48% (Grade 6) of the standards included in the study for each grade were considered DOK 1. These expectations emphasized use of standard algorithms to conduct calculations, recognition of particular Mathematics concepts, completion of problems where the solution path is evident, and reproduction of set procedures. Between 52% (Grade 6) and 81% (Grade 7) of the standards were considered a DOK level 2, targeting work involving conceptual understanding of Mathematics concepts, decision making, and/or making sense of Mathematics in context. Only three content standards were considered DOK 3. All of these standards were within the Grade 7 Statistics and Probability domain. These standards required students to evaluate probability models, design and use a simulation to find probabilities, and use appropriate samples of data to make inferences and predictions about an unknown characteristic. No standards were considered DOK 4. A DOK 4 expectation is one that is at least as complex as a DOK 3 but also requires extended time—days, weeks, or months—to complete. Although some components of these DOK 4 standards may be reasonably assessed by on-demand assessments, DOK 4 standards should not be expected to be fully assessed by an on-demand assessment. All of the expectations used in this study will be referred to as *standards* in this report.

Table 27. Content Expectations by Depth-of-Knowledge (DOK) Levels for West Virginia Mathematics Standards used in the Grades 3-8 Mathematics Alignment Analysis, 2019

Mathematics	Total Number of Expectations	DOK Level	Number of Standards by Level	Percent within Grade by Level
Grade 3	25	1 2	8 17	32 68
Grade 4	28	1 2	13 15	46 54
Grade 5	26	1 2	9 17	35 65
Grade 6	29	1 2	14 15	48 52
Grade 7	26	1 2 3	2 21 3	8 81 11
Grade 8	28	1 2	6 22	21 79

The West Virginia Mathematical Habits of Mind (MHMs) were also included in this study. The MHMs are not intended as stand-alone expectations but rather are intended to be integrated into the different content areas throughout the K-12 course of instruction. Panelists assigned a DOK to each, considering the intended uses of each MHM in the context of academic content. Most of the eight MHMs (75%) were considered to be DOK 3 and the remaining two (25%) were considered to be DOK 2.

Table 28. Mathematical Habits of Mind by Depth-of-Knowledge (DOK) Levels for West Virginia Mathematics Standards used in the Grades 3-8 Mathematics Alignment Analysis, 2019

Mathematics	Total Number of Expectations	DOK Level	Number of Standards by Level	Percent within Grade by Level
Grades 3 - 8	8	2 3	2 6	25 75

Test Forms

The West Virginia Mathematics assessment for Grades 3-8 consists of between 39-49 machine scorable items per test form. Items are one or two points each. No field test items were included in the analysis. The West Virginia Mathematics assessments include seven different item types, named and described in **Table 29**.

Table 29. Item Types for West Virginia Computer Adaptive General Mathematics Assessments for Grades 3-8

Response Type	Description
Equation (EQ)	Student uses a keypad with a variety of mathematical symbols to create a response. Responses can include numbers, fractions, expressions, inequalities, functions, and equations.
Editing Task Choice (ETC)	Student identifies an incorrect word or phrase and chooses the replacement from a number of options.
Grid (GI)	Student selects numbers, words, phrases, or images and uses the drag-and-drop feature to place them into a graphic. This item type may also require the student to use the point, line, or arrow tools to create a response on a graph.
Multiple Choice (MC)	Student selects one correct answer from four options.
Multiple Select (MS)	Student selects all correct answers from a number of options.
Table Input (TI)	Student types numeric values into a given table.
Table Match (MI)	Student checks a box to indicate if information from a column header matches information from a row.

WVGSA 2018-2019 Technical Report: Volume 1, p. 12

If no particular grade-level standard was targeted by a given assessment item, reviewers were instructed to code the item at a “higher” or more inclusive level, such as the domain level for Mathematics. This coding to a “generic standard” sometimes indicates that the item was inappropriate for a particular grade level (for example, the item might better match a standard from another grade level). If the item was grade-appropriate but a corresponding standard was not found, a generic coding may indicate that there is a part of the content within the standards that is being interpreted differently by different parties. Reviewers found that nearly all of the test items reasonably addressed specific standards. **Table 30**, on the next page, shows the items for each test form that a majority of reviewers coded to a generic standard. This table shows the generic standard to which the item was coded, the number of reviewers who coded the item to the generic standard, and the reason for the coding. No generic standards were used for 18 of the 20 Mathematics test forms reviewed. On just two of the 20 Mathematics test forms analyzed, a majority of reviewers mapped only one assessment item to a generic standard, indicating that the item did not directly target the content within any of the standards. Reviewers were required to write an explanation in the case of assigning an item to a generic standard. These notes were provided to the state and vendor in Appendix D, but redacted for public release. Items assigned to generic standards by a majority of panelists should be reviewed. It is possible that these items are inappropriately placed or coded in the item bank.

Table 30. Items Assigned to Generic Content Expectations by Assessment by a Majority of Reviewers, West Virginia Mathematics Alignment Analysis, 2019

Math Grade/Form	Generic Content Expectation	Item Number (# of Reviewers)	Reason
Grade 3 Batch 1	OA.M.3A	27(5)	Panelists commented that this item, which asks for a different representation of the problem with different factors, falls outside the scope of the Grade 3 standards.
Grade 4 Batch 2	NBT.NF.M.4A	9(5)	Panelists noted that the item uses numbers that exceed the digits limit within the Grade 4 standards.

Alignment Statistics and Findings for Mathematics Grades 3-8

Overall alignment results are summarized in **Table 31** on the next page and then detailed for each test form in the pages that follow. Based on typically accepted cutoffs for the four main alignment criteria considered in this study, all but one test form analyzed would be considered fully aligned with West Virginia College- and Career-Readiness Standards for Mathematics Grades 3-8. One of the 20 test forms analyzed had two reporting categories with weak or unmet DOK Consistency. These issues could be resolved with the revision or replacement of two items and the test form was therefore still considered acceptably aligned.

Table 31. Overall Alignment Findings for Grades 3-8 Mathematics Assessments with West Virginia College- and Career-Readiness Standards for Grades 3-8 Mathematics

Test Form	Independent Analysis Findings	Approx. Number of Items that Need Revision/Replacement for Full Alignment	Confirmatory Analysis Findings
Grade 3 B1	Fully Aligned	—	Fully meets
Grade 3 B2	Fully Aligned	—	NT*
Grade 3 B3	Fully Aligned	—	Fully meets
Grade 3 B4	Fully Aligned	—	NT
Grade 4 B1	Fully Aligned	—	Fully meets
Grade 4 B2	Fully Aligned	—	Fully meets
Grade 4 B4**	Fully Aligned	—	NT
Grade 5 B1	Fully Aligned	—	Fully meets
Grade 5 B2	Fully Aligned	—	Fully meets
Grade 5 B4**	Fully Aligned	—	NT
Grade 6 B1	Fully Aligned	—	Fully meets
Grade 6 B2 – ELEM GROUP	Fully Aligned	—	Fully meets
Grade 6 B2 – MS GROUP	Fully Aligned	—	Fully meets
Grade 6 B3	Fully Aligned	—	Fully meets
Grade 6 B4	Fully Aligned	—	Fully meets
Grade 7 B1	Acceptably Aligned	2 items	Fully meets
Grade 7 B2	Fully Aligned	—	Fully meets
Grade 7 B4	Fully Aligned	—	Fully meets
Grade 8 B1	Fully Aligned	—	Fully meets
Grade 8 B2	Fully Aligned	—	Fully meets
Grade 8 B4**	Fully Aligned	—	Fully meets

*Not Tested

**Coded by three reviewers

Three or more test events were analyzed per grade. Some items were used in more than one test event (repeated across test events). In these cases, reviewers' previous codings repopulated for efficiency in the item viewing interface (i.e. their "test booklet"). For each grade, between 88 and 114 unique items were reviewed in aggregate. This represented between 18% and 25% of the item pool for that grade. Overall, a total of 23% of the Grades 3-8 Mathematics item pool was analyzed in this study. The number and proportion of items reviewed per grade is detailed in **Table 32**. There was very little variation of results across test forms analyzed, suggesting that additional test events would likely yield similar results.

Table 32. Number of Unique Items Analyzed and Proportion of Grade Level Item Bank – West Virginia Grades 3-8 Mathematics Assessments

Grade	Number of unique items analyzed	Total Grade Level Operational Item Count	Proportion of grade level item bank
3	112	480	23%
4	88	496	18%
5	95	409	23%
6	114	473	24%
7	91	365	25%
8	94	415	23%
TOTAL (Math Grades 3-8)	594	2638	23%

Results by Test Form – Independent Analysis

The results of the analysis for each of the four alignment criteria are provided in **Tables 33 - 53** for each Mathematics test form for the corresponding reporting categories, which varies by grade. The approximate numbers of replaced or revised items necessary to meet minimum levels of alignment are provided for each test form. More detailed data on each of the criteria are given in **Appendix B**, in the first three tables for each test form.

In **Tables 33 - 53**, “YES,” indicates that an acceptable level was attained between the assessment and the reporting category on the criterion. “WEAK” indicates that the criterion was nearly met, within a margin that could simply be due to error or reasonable variation in reviewer coding. “NO” indicates that the criterion was not met by a noticeable margin—10% under an acceptable level for Depth-of-Knowledge Consistency, 10% under an acceptable level for Range-of-Knowledge Correspondence, and 0.1 under an index value of 0.7 for Balance-of-Representation. Categorical Concurrence is reported in average number of items. Depth-of-Knowledge Consistency is reported by the percent of items that were at or above the DOK of the corresponding standard. Range-of-Knowledge is reported as the percent of standards within each reporting category that were targeted by one or more items. Balance-of-Representation is an index value, ranging from 0-1.

Panelists were able to complete a review of at least three test forms per grade. If time allowed, three or more panelists reviewed a fourth form.

Mathematics Grade 3: Test Forms G3B1, G3B2, G3B3, and G3B4

Panelists reviewed all four Grade 3 test forms, one sampled from below proficiency (Form G3B1), two sampled from at proficiency (Forms G3B2 and G3B3), and one sampled from above proficiency (Form G3B4). All four Grade 3 test events analyzed met all alignment criteria for all reporting categories.

Table 33. Results for Mathematics Grade 3 Form B1 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 3

WV Grade 3 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.3 Operations and Algebraic Thinking (OA)	11.2	50%	76%	0.76	YES	YES	YES	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	18.7	78%	97%	0.84	YES	YES	YES	YES
MD.G.M.3 Measurement and Data and Geometry	9.3	57%	68%	0.84	YES	YES	YES	YES

*Number of items/points

Table 34. Results for Mathematics Grade 3 Form B2 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 3

WV Grade 3 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.3 Operations and Algebraic Thinking (OA)	12.7	57%	85%	0.72	YES	YES	YES	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	13.7	79%	86%	0.87	YES	YES	YES	YES
MD.G.M.3 Measurement and Data and Geometry	10.3	59%	76%	0.84	YES	YES	YES	YES

*Number of items/points

Table 35. Results for Mathematics Grade 3 Form B3 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 3

WV Grade 3 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.3 Operations and Algebraic Thinking (OA)	12.7	76%	70%	0.70	YES	YES	YES	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	15.8	82%	97%	0.86	YES	YES	YES	YES
MD.G.M.3 Measurement and Data and Geometry	9.2	69%	66%	0.83	YES	YES	YES	YES

*Number of items/points

Table 36. Results for Mathematics Grade 3 Form B4 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 3

WV Grade 3 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.3 Operations and Algebraic Thinking (OA)	10.3	74%	77%	0.77	YES	YES	YES	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	15.3	90%	97%	0.87	YES	YES	YES	YES
MD.G.M.3 Measurement and Data and Geometry	8.8	77%	76%	0.89	YES	YES	YES	YES

*Number of items/points

Mathematics Grade 4: Test Forms G4B1, G4B2, and G4B4

Panelists reviewed three test forms for Grade 4, one sampled from below proficiency (Form G4B1), one sampled from at proficiency (Form G4B2), and one sampled from above proficiency (Form G4B4). All Grade 4 test events analyzed met all alignment criteria for all reporting categories.

Table 37. Results for Mathematics Grade 4 Form B1 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 4

WV Grade 4 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.4 Operations and Algebraic Thinking (OA)	12.0	52%	96%	0.87	YES	YES	YES	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	15.3	69%	93%	0.85	YES	YES	YES	YES
MD.G.M.4 Measurement and Data and Geometry	9.0	74%	80%	0.89	YES	YES	YES	YES

*Number of items/points

Table 38. Results for Mathematics Grade 4 Form B2 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 4

WV Grade 4 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.4 Operations and Algebraic Thinking (OA)	9.2	59%	86%	0.85	YES	YES	YES	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	17.2	73%	83%	0.81	YES	YES	YES	YES
MD.G.M.4 Measurement and Data and Geometry	9.2	74%	80%	0.93	YES	YES	YES	YES

*Number of items/points

Table 39. Results for Mathematics Grade 4 Form B4 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 4 (Three Reviewers)

WV Grade 4 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.4 Operations and Algebraic Thinking (OA)	7.3	91%	86%	0.85	YES	YES	YES	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	17.3	62%	84%	0.83	YES	YES	YES	YES
MD.G.M.4 Measurement and Data and Geometry	10.3	92%	76%	0.82	YES	YES	YES	YES

*Number of items/points

Mathematics Grade 5: Test Forms G5B1, G5B2, and G5B4

All panelists reviewed two Grade 5 test forms, one sampled from below proficiency (Form G5B1), and one sampled from at proficiency (Form G5B2). Three panelists reviewed a third test event from sampled from above proficiency (Form G5B4). All Grade 5 test events analyzed met all alignment criteria for all reporting categories.

Table 40. Results for Mathematics Grade 5 Form B1 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 5

WV Grade 5 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.5 Operations and Algebraic Thinking (OA)	10.2	78%	100%	0.86	YES	YES	YES	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	15.8	51%	76%	0.83	YES	YES	YES	YES
MD.G.M.5 Measurement and Data and Geometry	10.0	56%	72%	0.84	YES	YES	YES	YES

*Number of items/points

Table 41. Results for Mathematics Grade 5 Form B2 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 5

WV Grade 5 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.5 Operations and Algebraic Thinking (OA)	10.7	75%	100%	0.84	YES	YES	YES	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	17.8	66%	67%	0.85	YES	YES	YES	YES
MD.G.M.5 Measurement and Data and Geometry	10.5	67%	77%	0.81	YES	YES	YES	YES

*Number of items/points

Table 42. Results for Mathematics Grade 5 Form B4 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 5 (Three Reviewers)

WV Grade 5 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
OA.M.5 Operations and Algebraic Thinking (OA)	8.0	66%	88%	0.79	YES	YES	YES	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	20.3	75%	69%	0.8	YES	YES	YES	YES
MD.G.M.5 Measurement and Data and Geometry	9.0	77%	77%	0.85	YES	YES	YES	YES

*Number of items/points

Mathematics Grade 6: Test Forms G6B1, G6B2, G6B3, and G6B4

Panelists reviewed four Grade 6 test forms: one sampled from below proficiency (Form G6B1), two sampled from at proficiency (Forms G6B2 and G6B3), and one sampled from above proficiency (G6B4). All test forms analyzed showed full alignment. Both the elementary and middle school panels analyzed Form G6B2 for the purpose of checking on and promoting inter-panel consistency in coding. The results of this analysis showed similar alignment findings for both panels. All Grade 6 test events analyzed met all alignment criteria for all reporting categories.

Table 43. Results for Mathematics Grade 6 Form B1 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 6

WV Grade 6 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	16.0	78%	93%	0.77	YES	YES	YES	YES
EE.M.6 Expressions and Equations	13.8	95%	75%	0.74	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	7.2	76%	66%	0.87	YES	YES	YES	YES

*Number of items/points

Table 44. Results for Mathematics Grade 6 Form B2 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 6 – Elementary Panel

WV Grade 6 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	15.2	77%	80%	0.80	YES	YES	YES	YES
EE.M.6 Expressions and Equations	11.7	83%	88%	0.81	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	7.8	91%	53%	0.80	YES	YES	YES	YES

*Number of items/points

Table 45. Results for Mathematics Grade 6 Form B2 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 6 – Middle School Panel

WV Grade 6 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	15.3	84%	83%	0.81	YES	YES	YES	YES
EE.M.6 Expressions and Equations	11.8	90%	92%	0.83	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	8.0	96%	55%	0.77	YES	YES	YES	YES

*Number of items/points

Table 46. Results for Mathematics Grade 6 Form B3 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 6

WV Grade 6 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	15.2	83%	89%	0.79	YES	YES	YES	YES
EE.M.6 Expressions and Equations	12.0	69%	72%	0.83	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	6.8	73%	70%	0.94	YES	YES	YES	YES

*Number of items/points

Table 47. Results for Mathematics Grade 6 Form B4 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 6

WV Grade 6 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	15.3	81%	86%	0.80	YES	YES	YES	YES
EE.M.6 Expressions and Equations	13.3	75%	92%	0.80	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	6.8	70%	74%	0.98	YES	YES	YES	YES

*Number of items/points

Mathematics Grade 7: Test Forms G7B1, G7B2, and G7B4

All panelists reviewed three Grade 7 test events, one sampled from below proficiency (Form G7B1), one sampled from at proficiency (Form G7B2), and one sampled from above proficiency (G7B4). Forms G7B2 and G7B4 met all alignment criteria for all reporting categories. Form G7B1 had unmet or weak DOK Consistency for two of the four reporting categories. For Form G7B1, two 1-point items (one that targets RP.NS.M.7 and one that targets G.M.7) would need to be revised or replaced to meet DOK Consistency. Both items would need to reflect complexity as expressed in the corresponding standard.

Table 48. Results for Mathematics Grade 7 Form B1 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 7

WV Grade 7 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	10.2	39%	83%	0.80	YES	NO	YES	YES
EE.M.7 Expressions and Equations	8.0	79%	100%	0.85	YES	YES	YES	YES
G.M.7 Geometry	7.8	48%	100%	0.84	YES	WEAK	YES	YES
SP.M.7 Statistics and Probability	9.0	59%	55%	0.87	YES	YES	YES	YES

*Number of items/points

Table 49. Results for Mathematics Grade 7 Form B2 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 7

WV Grade 7 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	11.3	69%	80%	0.81	YES	YES	YES	YES
EE.M.7 Expressions and Equations	9.7	85%	100%	0.9	YES	YES	YES	YES
G.M.7 Geometry	9.0	77%	100%	0.78	YES	YES	YES	YES
SP.M.7 Statistics and Probability	8.0	68%	60%	0.83	YES	YES	YES	YES

*Number of items/points

Table 50. Results for Mathematics Grade 7 Form B4 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 7 (by three panelists)

WV Grade 7 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	8.0	72%	62%	0.86	YES	YES	YES	YES
EE.M.7 Expressions and Equations	5.3	95%	56%	0.90	YES	YES	YES	YES
G.M.7 Geometry	7.8	95%	66%	0.77	YES	YES	YES	YES
SP.M.7 Statistics and Probability	6.8	66%	55%	0.90	YES	YES	YES	YES

*Number of items/points

Mathematics Grade 8: Test Forms G8B1, G8B2, and G8B4

All panelists reviewed two Grade 8 test forms, one sampled from below proficiency (Form G8B1), and one sampled from at proficiency (Form G8B2). Three panelists reviewed a third test event sampled from above proficiency (Form G8B4). All Grade 8 test events analyzed met all alignment criteria for all reporting categories.

Table 51. Results for Mathematics Grade 8 Form B1 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 8

WV Grade 8 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	12.0	69%	90%	0.83	YES	YES	YES	YES
F.M.8 Functions (F)	9.0	92%	100%	0.82	YES	YES	YES	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	15.0	72%	88%	0.82	YES	YES	YES	YES

*Number of items/points

Table 52. Results for Mathematics Grade 8 Form B2 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 8

WV Grade 8 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	12.2	79%	93%	0.84	YES	YES	YES	YES
F.M.8 Functions (F)	9.0	72%	80%	0.92	YES	YES	YES	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	14.8	53%	78%	0.78	YES	YES	YES	YES

*Number of items/points

Table 53. Results for Mathematics Grade 8 Form B4 and West Virginia College- and Career-Readiness Standards for Mathematics Grade 8 (Three Reviewers)

WV Grade 8 Math Reporting Categories	Alignment Statistics				Alignment Findings			
	CC*	DOK %	Range	Balance	CC	DOK	Range	Balance
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	9.3	68%	62%	0.88	YES	YES	YES	YES
F.M.8 Functions (F)	7.8	65%	90%	0.89	YES	YES	YES	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	10.8	82%	59%	0.87	YES	YES	YES	YES

*Number of items/points

Results by Test Form – Confirmatory Analysis

For the confirmatory analysis of each test form’s metadata, panelists considered the internally coded standard and responded to the question “*To what extent does the item assess the content (expectations) within the internally coded standard(s)?*” using the following categories:

- **EXACTLY** (Note that the item does NOT need to assess every aspect of a standard, but it needs to be a direct measurement of a central aspect of the standard. A correct response to the item allows for a direct inference about student knowledge/skills/abilities as expressed in the standard.)
- **PARTIALLY** (The item somewhat targets the expectations within the standard and it can be considered a majority match. A correct response to the item allows for some inference about student knowledge/skills/abilities as expressed in the standard.)
- **MINIMALLY** (The item only very minimally targets the expectations within the standard – and it can be considered only a minority match. A correct response to the item allows for very little or very indirect inference about student knowledge/skills/abilities as expressed in the standard.)
- **NOT AT ALL** (The item does not assess the expectations within the standard. No inference can be made about student knowledge/skills/abilities as expressed in the standard based on a correct response.)

For all test forms, panelists judged between 76% and 100% of items to be an exact measure of a central expectation within the internally coded standard. Panelists judged between 91% and 100% of the items on all test forms to be an exact or partial measure of the internally coded standard. In their notes, panelists provided their rationale(s) for any item that they thought only minimally addressed or did not at all address the internally coded standard. Results for the confirmatory analysis by test form is shown in **Table 54**. Item-level reviewer comments were provided in full to the state and to the test vendor but have been redacted from this report for public release.

The internally coded DOK uses a slightly different interpretation than the definitions for Webb's Depth-of-Knowledge. External reviewers coded fewer items as DOK 3 than are represented in the blueprints and internal metadata. Despite the differences, the independent analysis found that DOK Consistency was met for all reporting categories on 19 out of the 20 test forms reviewed.

Table 54. Results for West Virginia Math Grades 3-8 Test Forms Confirmatory Analysis

WV Math Test Events	Exact	Exact or Partially	Minimally and/or Not at All	Alignment Finding
Grade 3 B1	91%	100%	0%	Fully meets
Grade 3 B3	97%	100%	0%	Fully meets
Grade 4 B1	88%	100%	0%	Fully meets
Grade 4 B2	94%	100%	0%	Fully meets
Grade 5 B1	91%	100%	0%	Fully meets
Grade 5 B2	97%	100%	0%	Fully meets
Grade 6 B1	85%	100%	0%	Fully meets
Grade 6 B2 - ELEM PANEL	76%	94%	6%	Fully meets
Grade 6 B2 - MS PANEL	79%	91%	9%	Fully meets
Grade 6 B3	85%	97%	3%	Fully meets
Grade 6 B4	97%	97%	3%	Fully meets
Grade 7 B1	97%	100%	0%	Fully meets
Grade 7 B2	94%	100%	0%	Fully meets
Grade 7 B4*	100%	100%	0%	Fully meets
Grade 8 B1	94%	100%	0%	Fully meets
Grade 8 B2	94%	100%	0%	Fully meets
Grade 8 B4*	94%	100%	3%	Fully meets

*completed by three reviewers

Table 55 on the next page shows the specific items for which reviewers disagreed with internal coding by marking that the item only minimally targeted the internally coded standard or did not at all target the standard.

Table 55. Items Considered Miscoded by a Split or Majority of Panelists - West Virginia Mathematics Grades 3-8 Test Forms Confirmatory Analysis

Test Form	Item Number	Split/Majority Coding	Rationale
Grade 6 B2 - ELEM PANEL	#2 – 10199-2306	Not at all (6)	Panelists noted that the item did not address center, spread, of overall shape as specified in internally correlated standard and that the standard coding should be changed to Standard 6.28.
	#28 – REP10199- 1854	Minimally (5)	Panelists commented that the question asked students to solve a 2-step equation, which was a 7 th grade expectation, that no set of numbers was provided for substitution, and that, overall, students would get the answer by trial and error vs demonstration.
Grade 6 B2 - MS PANEL	#2 – 10199-2306	Not at all (5) Minimally (1)	See notes above for this item.
	#28 – REP10199- 1854	Minimally (4)	See notes above for this item.
	#29 – 10199-2125	Not at all (5) Minimally (1)	Panelists noted that this item targeted Standard 6.10 and is perhaps mislabeled.
Grade 6 B3	#15 – REP10199- 1854	Minimally (5)	Panelists commented that this was a grade 7 expectation.
Grade 6 B4	#17 – 10199-1858	Minimally (3)	Panelists commented that this was beyond the scope of grade 6 expectations.
Grade 8 B4*	#20 – 10199-2948	Not at all (25) Minimally (1)	Panelists commented that because the question stem provides a decimal approximation, the item does not target the intended standard. They noted that this could be resolved by removing the decimal approximation from the question stem.

Source of Challenge Issues and Reviewers' Comments Reviewers were instructed to document any Source-of-Challenge issue and to provide any other comments they may have about an item. A Source-of-Challenge is a technical issue with an item that can result in a student answering the item correctly or incorrectly for the wrong reason. Two items were identified by a majority of reviewers as having a Source-of-Challenge. These items and reviewer comments are provided in **Table 56** on the next page. All reviewer Source-of-Challenge comments should be reviewed as one person may have noticed something that others did not.

Reviewers also wrote notes about a number of items. Some notes indicate when only part of a particular standard was targeted by an assessment task. These notes also include general comments as well as suggestions for addressing issues identified. After coding each assessment form, reviewers were asked to respond to four debriefing questions. Item-level reviewer comments were provided in full to the state and to the test vendor but have been redacted from this report for public release.

Table 56. Items Flagged with Source-of-Challenge by a Majority of Reviewers, West Virginia Mathematics Alignment Analysis, 2019

Math Grade/Form	Item Number	Reason
Grade 3 Batch 1	23	Panelists commented that the item references a rectangle but that it is not clear which rectangle is being referenced.
Grade 7 Batch 1 (REP in G7B4)	27 (28)	Panelists flagged an editorial error in this problem. They noted that the expression given is for total money but that the item stem asks for total books.

Reliability among Reviewers

Reviewers engaged in some adjudication of their data after all reviewers finished their coding for an assessment. After the first two or three test events, broad differences in coding were limited and panels conducted very little or no adjudication. These discussions were used to identify any mistakes in coding. Reviewers were not required to change their coding after discussion unless they found a compelling reason. The agreement statistics shown in **Table 57**, on the following page, were computed after adjudication.

The overall intraclass correlation among the math reviewers' assignment of DOK levels to items was reasonable or high for nearly all analyses (higher than 0.80). For five test forms, the intraclass correlation for DOK was below the desired value of 0.80 (G4B2: 0.60, G5B2: 0.77, G5B4: 0.77, G7B4: 0.73, and G8B4: 0.73). Three of these test forms were coded by only three of the reviewers and is likely the reason for a slightly lowered agreement (0.73 or 0.77). If all of the reviewers had reviewed these forms, it is likely the intraclass correlation would be higher, most likely above 0.80. The intraclass correlation for the coding of DOK on Form 4B2 of 0.60 is lower than desired. Results pertaining to DOK for this form should take into consideration that variation among reviewers in coding the DOK of items was larger than usual in most alignment studies. The intraclass correlation of 0.77 for Form 5B2 is only slightly below the desired value of 0.80 and is not considered a major issue (**Table 57**).

A pairwise comparison was used to determine the degree of reliability of reviewers coding at the reporting category level and the standard level. The pairwise comparison was computed by considering for every item the coding assigned by each reviewer compared to the coding by each of the other five reviewers. For example, for six reviewers a total of 15 comparisons are computed for each item. For most alignment studies, the objective pairwise agreement is higher than 0.6. The pairwise agreement for assigning objectives to items met this benchmark for all test forms analyzed except for one (G5B4, which was coded by only three panelists) and was very high compared to most studies. For coding to the level of reporting category, a pairwise agreement of 0.90

is desired. For all but one test form, pairwise agreement for reporting category was close to or above 0.90. Form G5B4 had a low pairwise agreement (0.69) for reporting category. The conclusions for Form G5B4 should take into consideration that reviewers had some variation in assigning items to standards and reporting categories.

Table 57. Intraclass and Pairwise Comparisons, West Virginia Grades 3-8 Mathematics Assessments with College- and Career-Readiness Standards for Mathematics Grades 3-8

Test Form	Intraclass Correlation (DOK)	Pairwise Comparison (DOK)	Pairwise Comparison (Reporting Category)	Pairwise Comparison (Objectives)
Grade 3 B1	0.94	0.79	0.91	0.77
Grade 3 B2	0.80	0.64	0.91	0.74
Grade 3 B3	0.88	0.73	0.90	0.73
Grade 3 B4	0.89	0.75	0.87	0.74
Grade 4 B1	0.91	0.76	0.90	0.73
Grade 4 B2	0.60	0.56	0.86	0.66
Grade 4 B4*	0.87	0.75	0.88	0.75
Grade 5 B1	0.83	0.71	0.96	0.84
Grade 5 B2	0.77	0.62	0.97	0.84
Grade 5 B4*	0.77	0.65	0.69	0.45
Grade 6 B1	0.93	0.80	0.97	0.82
Grade 6 B2 - ELEM PANEL	0.93	0.79	0.92	0.72
Grade 6 B2 - MS PANEL	0.95	0.83	0.90	0.82
Grade 6 B3	0.85	0.72	0.96	0.85
Grade 6B4	0.91	0.73	0.93	0.78
Grade 7B1	0.90	0.74	0.96	0.88
Grade 7B2	0.85	0.72	0.97	0.91
Grade 7B4*	0.73	0.72	1.00	0.89
Grade 8B1	0.90	0.77	0.98	0.96
Grade 8B2	0.89	0.70	0.99	0.96
Grade 8B4*	0.73	0.63	0.96	0.91

Conclusion

An alignment analysis was conducted during summer 2019 to provide information about the degree of alignment of the computer adaptive West Virginia statewide general summative assessments of Grades 3-8 ELA and Grades 3-8 Mathematics with the corresponding West Virginia College- and Career-Readiness Standards, as pertains to fulfilling requirements as stated in Federal statute.

The overall alignment analysis conducted for the West Virginia Department of Education incorporated information from assessment framework documents including blueprints and narrative information about the item selection algorithm, from statistical analyses provided within assessment technical reports of aggregate data from all test events administered in the state, and from a two-part content analysis. Consideration of framework and statistical documentation were conducted remotely. These components attended to the assessment design and the test developer's claims. The other study component involved a two-part content analysis. Content analyses were conducted by four (4) panels of six (6) expert educators during an in-person meeting. The overall study was designed to allow for the potential to craft a logic argument for the capacity of alignment of all test events generated by the West Virginia statewide general summative computer adaptive test (CAT) assessments for Grades 3-8 ELA and Grades 3-8 Mathematics with the corresponding West Virginia College- and Career-Readiness Standards, as appropriate, based on results.

The in-person content alignment analysis was designed to determine 1) the degree to which the simulated test events that were used in the analysis were aligned with the West Virginia College- and Career-Readiness Standards for ELA and Mathematics and 2) the accuracy and appropriateness of the vendor's internal metadata. The vendor's internal metadata specified the vendor's intended assessment target for each item, including the standard and DOK level of the item. Because the West Virginia ELA and Mathematics assessments are computer adaptive, the accuracy of the internal metadata is critical; appropriate selection of items for use in test events per test blueprints/algorithm is dependent on the internal metadata, particularly the coding for targeted standards.

The results of the in-person content analyses were compared with framework and statistical documentation to make an inference about the capacity of the overall test program to generate aligned test events for each subject and grade. A summary of overall Grades 3-8 Mathematics and ELA assessment program capacity to generate assessments aligned with corresponding grade-level standards is summarized by alignment criterion on the following page.

Categorical Concurrence: Both ELA and Mathematics assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for all reporting categories.

DOK Consistency: ELA assessments for Grades 3-6 and Mathematics assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for all reporting categories. ELA assessments for Grades 7-8 need slight adjustments to meet this criterion for the Reading Literary Text and Reading Informational Text reporting categories.

Range-of-Knowledge: Mathematics assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for all reporting categories. ELA assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for Literary Reading and Informational Reading reporting categories. ELA assessments for Grades 3-8 have the capacity to fully meet or very nearly meet this alignment criterion for the Writing and Language reporting categories, with consideration for comparison against an appropriately condensed set of standards.

Balance of Representation: Mathematics assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for all reporting categories. ELA assessments for Grades 3-8 have the capacity to fully meet this alignment criterion for Literary Reading and Informational Reading reporting categories. Any weak Balance for the Writing and Language reporting category on the ELA assessments for Grades 3-8 can be resolved along with Range and is not considered an alignment issue on its own.

A summary of overall Grades 3-8 Mathematics and Grades 3-8 ELA assessment program capacity to generate assessments aligned with corresponding grade-level standards is summarized below, along with supporting evidence.

Overall, the results of a two-part content alignment analysis along with a consideration of blueprints, information about the item selection algorithm, and aggregate data from administered test events as reported in the assessment technical report suggest that the West Virginia statewide general summative computer adaptive test program for Grades 3-8 Mathematics and Grades 3-6 ELA has the capacity to generate aligned test events across a range of proficiency. The evidence to support this claim includes:

- A confirmatory content analysis of items on each of 11 test events for ELA Grades 3-6 and 19 test events for Mathematics Grades 3-8 suggests that the items are appropriate as relates to intended claims and inferences about the targeted content (standard) based on comparison of internal metadata with independent external analysis.
- According to the aggregate data for all administered test events as shown in the 2018-2019 West Virginia Technical Report, the test blueprints and algorithm are generating test events as intended; with the exception of known and uncommon circumstances, all tests delivered met blueprint specifications.
- Further, an independent analysis of test events sampled from a range of proficiency levels for Grades 3-8 Mathematics and Grades 3-6 ELA were found to fully or acceptably meet minimum alignment criteria according to typically used cutoffs for Categorical Concurrence, DOK Consistency, Range of Knowledge Correspondence, and Balance of Representation.

For all test events analyzed for ELA Grades 7 and 8, all reporting categories met most alignment criteria but all test events were found to need adjustments to meet the minimum cutoffs for overall alignment. To be fully aligned, each test event required the revision or replacement of around seven items per test form to address weak or unmet DOK consistency for both reading reporting categories. Revisions or amendments to the Grade 7 and 8 item bank could ensure that there are a sufficient proportion of DOK 3 items corresponding to DOK 3 standards. WVDE may also wish to review the specifications for distribution of DOK (“DOK Ranges”) given in the test blueprints in relation to the distribution of DOK in the corresponding grade-level standards. The test blueprint specifies around 25% of the items on any ELA test form are expected to be DOK 3. In contrast, around 70% of Grades 7 and 8 standards were considered DOK 3. As noted previously, the external reviewers did not always agree with the internally coded DOK, so other factors could also be in play such as differences in training or in the use of references (Webb-sourced vs alternate interpretations). Because DOK consistency for ELA Grades 7 and 8 was the only significant alignment gap, the assessment program could be expected to generate acceptably aligned test events for these grades with appropriate revisions in place.

WVDE may also wish to review the blueprint specifications related to targeting Mathematical Habits of Mind (MHM) as panelists found very few items they thought truly elicited the intent of the MHMs in the summative assessment items. This is not considered an alignment issue, as MHMs are not necessarily appropriately measured in a single snapshot, as on a summative assessment. However, WVDE may wish to reconsider the claims about MHM sub-score reporting. For example, although the assessments do not elicit the full intent of the MHMs, the items may contribute to student engagement with some aspects of these MHMs.

References

- West Virginia Department of Education. (2018) West Virginia Statewide Assessment Program 2017-2018 Annual Technical Report: Volume 1
- West Virginia Department of Education. (2018) West Virginia Statewide Assessment Program 2017-2018 Annual Technical Report: Volume 2
- Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25(1), 47-55.
- Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literary standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III*. Mahwah, NJ: Lawrence Erlbaum.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in Mathematics and Mathematics education*. Council of Chief State School Officers and National Institute for Mathematics Education Research Monograph No. 6. Madison: University of Wisconsin, Wisconsin Center for Education Research.
- Webb, N. L. (1999). Alignment of science and mathematics standards and assessments in four states. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 18. Madison, WI: University of Wisconsin.
- Webb, N. L. (2003). Alignment study in language arts, mathematics, science, and social studies of state standards and assessments in four states. Washington, D. C.: Council of Chief State School officers.
- Wise, S. L., Kingsbury, G. G., and Webb, N. L. (2015). Evaluating content alignment in computerized adaptive testing. *Educational Measurement: Issues and Practices*, Winter, 34, 4, pp. 41-48.

Appendix A

Group Consensus DOK Values for West Virginia College- and Career- Readiness Standards for English Language Arts Grades 3-8

2019

	West Virginia College and Career Readiness Standards for Grade 3 ELA	
Standard	Description	DOK
LR.3	Grade 3 Literary Reading	
LR.3.KID	Key Ideas and Details	
ELA.3.1	Ask and answer questions to demonstrate understanding of a literary text, referring explicitly to the text as the basis for the answers.	2
ELA.3.2	Recount stories, including fables, folktales, and myths from diverse cultures; determine the central message, lesson, or moral and explain how it is conveyed through key details in the literary text.	2
ELA.3.3	Describe characters in a literary story (e.g., their traits, motivations, or feelings) and explain how their actions contribute to the sequence of events.	3
LR.3.CS	Craft and Structure	
ELA.3.7	Determine the meaning of words and phrases as they are used in a literary text, distinguishing literal from nonliteral language.	2
ELA.3.8	Refer to parts of stories, dramas, and poems when writing or speaking about a literary text, using terms such as chapter, scene, and stanza; describe how each successive part builds on earlier sections.	2
ELA.3.9	Distinguish one's own point of view from that of the narrator or those of the characters in a literary text.	2
LR.3.IKI	Integration of Knowledge and Ideas	
ELA.3.13	Explain how specific aspects of a literary text's illustrations contribute to what is conveyed by the words in a story (e.g., create mood or emphasize aspects of a character or setting).	3
ELA.3.14	Compare and contrast the themes, settings, and plots of literary stories written by the same author about the same or similar characters (e.g., in books from a series).	3

IR.3	Grade 3 Informational Reading	
IR.3.KID	Key Ideas and Details	
ELA.3.4	Ask and answer questions to demonstrate understanding of an informational text, referring explicitly to the text as the basis for the answers.	2
ELA.3.5	Determine the main idea of an informational text; recount the key details and explain how they support the main idea.	2
ELA.3.6	Describe the relationship between a series of historical events, scientific ideas or concepts, or steps in technical procedures in an informational text, using language that pertains to time, sequence, and cause/effect.	3
IR.3.CS	Craft and Structure	
ELA.3.10	Determine the meaning of general academic and domain-specific words and phrases in an informational text relevant to a grade 3 topic or subject area.	2
ELA.3.11	Use informational text features and search tools (e.g., key words, sidebars, and hyperlinks) to locate information relevant to a given topic efficiently.	2
ELA.3.12	Distinguish one's own point of view from that of the author of an informational text.	3
IR.3.IKI	Integration of Knowledge and Ideas	
ELA.3.15	Use information gained from illustrations (e.g., maps or photographs) and the words in an informational text to demonstrate understanding of the text (e.g., where, when, why, and how key events occur).	3
ELA.3.16	Describe the logical connection between particular sentences and paragraphs in an informational text (e.g., comparison, cause/effect, or first/second/third in a sequence).	2
ELA.3.17	Compare and contrast the most important points and key details presented in two informational texts on the same topic.	3

WL.3	Grade 3 Writing and Language	
WL.3.TTP	Writing - Text Types and Purposes	
ELA.3.20	Write opinion pieces on topics or texts, supporting a point of view with reasons. a. Introduce the topic or text they are writing about, state an opinion, and create an organizational structure that lists reasons. b. Provide reasons that support the opinion. c. Use linking words and phrases (e.g., because, therefore, since, for example) to connect opinion and reasons. d. Provide a concluding statement or section.	3
ELA.3.21	Write informative/explanatory texts to examine a topic and convey ideas and information clearly. a. Introduce a topic and group related information together; include illustrations when useful to aiding comprehension. b. Develop the topic with facts, definitions, and details. c. Use linking words and phrases (e.g., also, another, and, more, or but) to connect ideas within categories of information. d. Provide a concluding statement or section.	3
ELA.3.22	Write narratives to develop real or imagined experiences or events using effective technique, descriptive details, and clear event sequences. a. Establish a situation and introduce a narrator and/or characters; organize an event sequence that unfolds naturally. b. Use dialogue and descriptions of actions, thoughts, and feelings to develop experiences and events or show the response of characters to situations. c. Use transitional words and phrases to signal event order. d. Provide a sense of closure.	3
WL.3.PDW	Writing - Production and Distribution of Writing	
ELA.3.23	With guidance and support from adults, produce writing in which the development and organization are appropriate to task and purpose. (Grade-specific expectations for writing types are defined in Text Types and Purposes.)	3
ELA.3.24	With guidance and support from peers and adults, develop and strengthen writing as needed by planning, revising, and editing. (Editing for conventions should demonstrate command of Language standards up to and including grade 3.)	3
ELA.3.25	With guidance and support from adults, use technology to produce and publish writing (using keyboarding skills) as well as to interact and collaborate with others.	3

WL.3.RBPK	Writing - Research to Build and Present Knowledge	
ELA.3.26	Conduct short research projects that build knowledge about a topic.	3
ELA.3.27	Recall information from experiences or gather information from print and digital sources; take brief notes on sources and sort evidence into provided categories.	2
WL.3.RW	Writing - Range of Writing	
ELA.3.29	Write routinely over extended time frames (time for research, reflection, and revision) and shorter time frames (a single sitting or a day or two) for a range of discipline-specific tasks, purposes, and audiences.	4
WL.3.CSE	Language - Conventions of Standard English	
ELA.3.36	Demonstrate command of the conventions of Standard English grammar and usage when writing or speaking. a. Explain the function of nouns, pronouns, verbs, adjectives, and adverbs in general and their functions in particular sentences. b. Form and use regular and irregular plural nouns. c. Use abstract nouns (e.g., childhood). d. Form and use regular and irregular verbs. e. Form and use the simple (e.g., I walked; I walk; I will walk) verb tenses. f. Ensure subject-verb and pronoun-antecedent agreement. g. Form and use comparative and superlative adjectives and adverbs, and choose between them depending on what is to be modified. h. Use coordinating and subordinating conjunctions. i. Produce simple, compound, and complex sentences.	1
ELA.3.37	Demonstrate command of the conventions of Standard English capitalization, punctuation, and spelling when writing. a. Capitalize appropriate words in titles. b. Use commas in addresses. c. Use commas and quotation marks in dialogue. d. Form and use possessives. e. Use conventional spelling for high-frequency and other studied words and for adding suffixes to base words (e.g., sitting, smiled, cries, or happiness). f. Use spelling patterns and generalizations (e.g., word families, position-based spellings, syllable patterns, ending rules, and meaningful word parts) in writing words. g. Consult reference materials, including beginning dictionaries, as needed to check and correct spellings.	1

WL.3.KL	Language - Knowledge of Language	
ELA.3.38	Use knowledge of language and its conventions when writing, speaking, reading, or listening. a. Choose words and phrases for effect. b. Recognize and observe differences between the conventions of spoken and written Standard English.	2
WL.3.VAU	Language- Vocabulary Acquisition and Use	
ELA.3.39	Determine or clarify the meaning of unknown and multiple-meaning word and phrases based on grade 3 reading and content, choosing flexibly from a range of strategies. a. Use sentence-level context as a clue to the meaning of a word or phrase. b. Determine the meaning of the new word formed when a known affix is added to a known word (e.g., agreeable/disagreeable, comfortable/uncomfortable, care/careless, heat/preheat). c. Use a known root word as a clue to the meaning of an unknown word with the same root (e.g., company, companion). d. Use glossaries or beginning dictionaries, both print and digital, to determine or clarify the precise meaning of key words and phrases.	2
ELA.3.40	Demonstrate understanding of word relationships and nuances in word meanings. a. Distinguish the literal and nonliteral meanings of words and phrases in context (e.g., take steps). b. Identify real-life connections between words and their use (e.g., describe people who are friendly or helpful). c. Distinguish shades of meaning among related words that describe states of mind or degrees of certainty (e.g., knew, believed, suspected, heard, and wondered).	2
ELA.3.41	Acquire and use accurately grade-appropriate conversational, general academic, and domain-specific words and phrases, including those that signal spatial and transitional relationships (e.g., After dinner that night we went looking for them).	2

	West Virginia College and Career Readiness Standards for Grade 4 ELA	
Standard	Description	DOK
LR.4	Grade 4 Literary Reading	
LR.4.KID	Key Ideas and Details	
ELA.4.1	Refer to details and examples in a literary text when explaining what the text says explicitly and when drawing inferences from the text.	2
ELA.4.2	Determine a theme of a story, drama, or poem from details in the literary text; summarize the text.	2
ELA.4.3	Describe in depth a character, setting, or event in a story or drama, drawing on specific details in the literary text (e.g., a character's thoughts, words, or actions).	2
LR.4.CS	Craft and Structure	
ELA.4.7	Determine the meaning of words and phrases as they are used in a literary text, including those that allude to significant characters found in mythology (e.g., Herculean).	2
ELA.4.8	Explain major differences between poems, drama, and prose; refer to the structural elements of poems (e.g., verse, rhythm, meter) and drama (e.g., casts of characters, settings, descriptions, dialogue, and stage directions) when writing or speaking about a literary text.	2
ELA.4.9	Compare and contrast the point of view from which different literary texts are narrated, including the difference between first-and third-person narrations.	2
LR.4.IKI	Integration of Knowledge and Ideas	
ELA.4.13	Make connections between the text of a story or drama and a visual or oral presentation of the literary text, identifying where specific descriptions and directions in the text are reflected in the visual or oral presentation.	3
ELA.4.14	Compare and contrast the treatment of similar themes and topics (e.g., opposition of good and evil) and patterns of events (e.g., the quest) in stories, myths, traditional literature and literary text from different cultures.	3

IR.4	Grade 4 Informational Reading	
IR.4.KID	Key Ideas and Details	
ELA.4.4	Refer to details and examples in an informational text when explaining what the text says explicitly and when drawing inferences from the text.	2
ELA.4.5	Determine the main idea of an informational text and explain how it is supported by key details; summarize the text.	2
ELA.4.6	Explain events, procedures, ideas, or concepts in a historical, scientific, or technical text, including what happened and why, based on specific information in the informational text.	2
IR.4.CS	Craft and Structure	
ELA.4.10	Determine the meaning of general academic and domain-specific words or phrases in an informational text relevant to a grade 4 topic or subject area.	2
ELA.4.11	Describe the overall structure (e.g., chronology, comparison, cause/effect, or problem/solution) of events, ideas, concepts, or information in all or part of an informational text.	2
ELA.4.12	Compare and contrast a firsthand and secondhand account of the same event or topic; describe the differences in focus, and the information provided in these informational texts.	3
IR.4.IKI	Integration of Knowledge and Ideas	
ELA.4.15	Interpret information presented visually, orally, or quantitatively (e.g., in charts, graphs, diagrams, time lines, animations, or interactive elements on web pages) and explain how the information contributes to an understanding of the informational text in which it appears.	3
ELA.4.16	Explain how an author uses reasons and evidence to support particular points in an informational text.	3
ELA.4.17	Integrate information from two informational texts on the same topic in order to write or speak about the subject knowledgeably.	3

WL.4	Grade 4 Writing and Language	
WL.4.TTP	Writing - Text Types and Purposes	
ELA.4.20	Write opinion pieces on topics or texts, supporting a point of view with reasons and information. a. Introduce a topic or text clearly, state an opinion, and create an organizational structure in which related ideas are grouped to support the writer's purpose. b. Provide reasons that are supported by facts and details. c. Link opinion and reasons using words and phrases (e.g., for instance, in order to, or in addition). d. Provide a concluding statement or section related to the opinion presented.	3
ELA.4.21	Write informative/explanatory texts to examine a topic and convey ideas and information clearly. a. Introduce a topic clearly and group related information in paragraphs and sections; include formatting (e.g., headings), illustrations, and multimedia when useful to aiding comprehension. b. Develop the topic with facts, definitions, concrete details, quotations, or other information and examples related to the topic. c. Link ideas within categories of information using words and phrases (e.g., another, for example, also, or because). d. Use precise language and domain-specific vocabulary to inform about or explain the topic. e. Provide a concluding statement or section related to the information or explanation presented.	3
ELA.4.22	Write narratives to develop real or imagined experiences or events using effective technique, descriptive details, and clear event sequences. a. Orient the reader by establishing a situation and introduce a narrator and/or characters; organize an event sequence that unfolds naturally. b. Use dialogue and description to develop experiences and events or show the responses of characters to situations. c. Use a variety of transitional words and phrases to manage the sequence of events. d. Use concrete words and phrases and sensory details to convey experiences and events precisely. e. Provide a conclusion that follows from the narrated experiences or events.	3
WL.4.PDW	Writing - Production and Distribution of Writing	
ELA.4.23	Produce clear and coherent writing in which the development and organization are appropriate to task, purpose, and audience. (Grade-specific expectations for writing types are defined in Text Types and Purposes).	3
ELA.4.24	With guidance and support from peers and adults, develop and strengthen writing as needed by planning, revising, and editing. (Editing for conventions should demonstrate command of Language standards up to and including grade 4).	3
ELA.4.25	With some guidance and support from adults, use technology, including the Internet, to produce and publish writing as well as to interact and collaborate with others; demonstrate sufficient command of keyboarding skills.	2

WL.4.RBPK	Writing - Research to Build and Present Knowledge	
ELA.4.26	Conduct short research projects that build knowledge through investigation of different aspects of a topic.	3
ELA.4.27	Recall relevant information from experiences or gather relevant information from print and digital sources; take notes and categorize information and provide a list of sources.	2
ELA.4.28	Draw evidence from literary or informational texts to support analysis, reflection, and research. a. Apply grade 4 Reading standards to literature (e.g., “describe in depth a character, setting, or event in a story or drama, drawing on specific details in the text [e.g. a character’s thoughts, words, or actions].”). b. Apply grade 4 Reading standards to informational texts (e.g., “explain how an author uses reasons and evidence to support particular points in a text.”).	3
WL.4.RW	Writing - Range of Writing	
ELA.4.29	Write routinely over extended time frames (time for research, reflection, and revision) and shorter time frames (a single sitting or a day or two) for a range of discipline-specific tasks, purposes, and audiences.	4
WL.4.CSE	Language - Conventions of Standard English	
ELA.4.36	Demonstrate command of the conventions of Standard English grammar and usage when writing or speaking. a. Use relative pronouns (who, whose, whom, which, that) and relative adverbs (where, when, why). b. Form and use the progressive verb tenses (e.g., I was walking; I am walking; I will be walking). c. Use modal auxiliaries (e.g., can, may, must) to convey various conditions. d. Order adjectives within sentences according to conventional patterns (e.g., a small red bag rather than a red small bag). e. Form and use prepositional phrases. f. Produce complete sentences, recognizing and correcting inappropriate fragments and run-ons. g. Correctly use frequently confused words (e.g., to, too, two; there, their).	2
ELA.4.37	Demonstrate command of the conventions of Standard English capitalization, punctuation, and spelling when writing. a. Use correct capitalization. b. Use commas and quotation marks to mark direct speech and quotations from a text. c. Use a comma before a coordinating conjunction in a compound sentence. d. Spell grade-appropriate words correctly, consulting references as needed.	1

WL.4.KL	Language - Knowledge of Language	
ELA.4.38	Use knowledge of language and its conventions when writing, speaking, reading, or listening. a. Choose words and phrases to convey ideas precisely. b. Choose punctuation for effect. c. Differentiate between contexts that call for formal English (e.g., presenting ideas) and situations where informal discourse is appropriate (e.g., small-group discussion).	2
WL.4.VAU	Language - Vocabulary Acquisition and Use	
ELA.4.39	Determine or clarify the meaning of unknown and multiple-meaning words and phrases based on grade 4 reading and content, choosing flexibly from a range of strategies. a. Use context (e.g., definitions, examples, or restatements in text) as a clue to the meaning of a word or phrase. b. Use common, grade-appropriate Greek and Latin affixes and roots as clues to the meaning of a word (e.g., telegraph, photograph, autograph). c. Consult reference materials (e.g., dictionaries, glossaries, thesauruses), both print and digital, to find the pronunciation and determine or clarify the precise meaning of key words and phrases.	2
ELA.4.40	Demonstrate understanding of figurative language, word relationships, and nuances in word meanings. a. Explain the meaning of simple similes and metaphors (e.g., as pretty as a picture) in context. b. Recognize and explain the meaning of common idioms, adages, and proverbs. c. Demonstrate understanding of words by relating them to their opposites (antonyms) and to words with similar but not identical meanings (synonyms).	3
ELA.4.41	Acquire and accurately use grade-appropriate general academic and domain-specific words and phrases, including those that signal precise actions, emotions, or states of being (e.g., quizzed, whined, stammered) and that are basic to a particular topic (e.g., wildlife, conservation, and endangered when discussing animal preservation).	2

	West Virginia College and Career Readiness Standards for Grade 5 ELA	
Standard	Description	DOK
LR.5	Grade 5 Literary Reading	
LR.5.KID	Key Ideas and Details	
ELA.5.1	Quote accurately from a literary text when explaining what the text says explicitly and when drawing inferences from the text.	2
ELA.5.2	Determine a theme of a story, drama, or poem from details in a literary text, including how characters in a story or drama respond to challenges, how the speaker in a poem reflects upon a topic; summarize the text.	3
ELA.5.3	Compare and contrast two or more characters, settings, or events in a story or drama, drawing on specific details in the literary text (e.g., how characters interact).	3
LR.5.CS	Craft and Structure	
ELA.5.7	Determine the meaning of words and phrases as they are used in a literary text, including figurative language such as metaphors and similes.	2
ELA.5.8	Explain how a series of chapters, scenes, or stanzas fits together in a literary text to provide the overall structure of a particular story, drama, or poem.	2
ELA.5.9	Describe how a narrator's or speaker's point of view influences how events are described in a literary text.	2
LR.5.IKI	Integration of Knowledge and Ideas	
ELA.5.13	Analyze how visual and multimedia elements contribute to the meaning, tone, or beauty of a literary text (e.g., graphic novel, multimedia presentation of fiction, folktale, myth, and/or poem).	3
ELA.5.14	Compare and contrast stories in literary texts of the same genre (e.g., mysteries and adventure stories) on their approaches to similar themes and topics.	3

IR.5	Grade 5 Informational Reading	
IR.5.KID	Key Ideas and Details	
ELA.5.4	Quote accurately from an informational text when explaining what the text says explicitly and when drawing inferences from the text.	2
ELA.5.5	Determine two or more main ideas of an informational text and explain how they are supported by key details; summarize the text.	2
ELA.5.6	Using an informational text, explain the relationships or interactions between two or more individuals, events, ideas, or concepts in a historical, scientific, or technical text, based on specific information in the text.	2
IR.5.CS	Craft and Structure.	
ELA.5.10	Determine the meaning of general academic and domain-specific words and phrases in an informational text relevant to a grade 5 topic or subject area.	2
ELA.5.11	Compare and contrast the overall structure (e.g., chronology, comparison, cause/effect, and problem/solution) of events, ideas, concepts, or information in two or more informational texts.	2
ELA.5.12	Analyze multiple accounts of the same event or topic, noting important similarities and differences in the point of view they represent in informational texts.	3
IR.5.IKI	Integration of Knowledge and Ideas	
ELA.5.15	Draw on information from multiple print or digital informational sources, demonstrating the ability to locate an answer to a question quickly or to solve a problem efficiently.	3
ELA.5.16	Explain how an author uses reasons and evidence to support particular points in an informational text, identifying which reasons and evidence support which point(s).	3
ELA.5.17	Integrate information from several informational texts on the same topic in order to write or speak about the subject knowledgeably.	3

WL.5	Grade 5 Writing and Language	
WL.5.TTP	Writing - Text Types and Purposes	
ELA.5.20	Write opinion pieces on topics or texts, supporting a point of view with reasons and information. - Introduce a topic or text clearly, state an opinion, and create an organizational structure in which ideas are logically grouped to support the writer's purpose. - Provide logically ordered reasons that are supported by facts and details. - Link opinion and reasons using words, phrases, and clauses (e.g., consequently and specifically). - Provide a concluding statement or section related to the opinion presented.	3
ELA.5.21	Write informative/explanatory texts to examine a topic and convey ideas and information clearly. - Introduce a topic clearly, provide a general observation, and focus and group related information logically; include formatting (e.g., headings), illustrations, and multimedia when useful to aiding comprehension. - Develop the topic with facts, definitions, concrete details, quotations, or other information and examples related to the topic. - Link ideas within and across categories of information using words, phrases, and clauses (e.g., in contrast, especially). - Use precise language and domain-specific vocabulary to inform about or explain the topic. - Provide a concluding statement or section related to the information or explanation presented.	3
ELA.5.22	Write narratives to develop real or imagined experiences or events using effective technique, descriptive details, and clear event sequences. - Orient the reader by establishing a situation and introducing a narrator and/or characters; organize an event sequence that unfolds naturally. - Use narrative techniques, such as dialogue, description, and pacing, to develop experiences and events or show the responses of characters to situations. - Use a variety of transitional words, phrases, and clauses to manage the sequence of events. - Use concrete words and phrases and sensory details to convey experiences and events precisely. - Provide a conclusion that follows from the narrated experiences or events.	3
WL.5.PDW	Writing - Production and Distribution of Writing	
ELA.5.23	Produce clear and coherent writing in which the development and organization are appropriate to task, purpose, and audience. (Grade-specific expectations for writing types are defined in Text Types and Purposes.)	3
ELA.5.24	With guidance and support from peers and adults, develop and strengthen writing as needed by planning, revising, editing, rewriting, or trying a new approach. (Editing for conventions should demonstrate command of Language standards up to and including grade 5.)	3
ELA.5.25	With some guidance and support from adults, use technology, including the Internet, to produce and publish writing as well as to interact and collaborate with others; demonstrate sufficient command of keyboarding skills to type accurately.	2

WL.5.RBPK	Writing - Research to Build and Present Knowledge	
ELA.5.26	Conduct short research projects that use several sources to build knowledge through investigation of different aspects of a topic.	3
ELA.5.27	Recall relevant information from experiences or gather relevant information from print and digital sources; summarize or paraphrase information in notes and finished work, and provide a list of sources.	2
ELA.5.28	Draw evidence from literary or informational texts to support analysis, reflection, and research. a. Apply grade 5 Reading standards to literature (e.g., “compare and contrast two or more characters, settings, or events in a story or drama, drawing on specific details in the text [e.g. how characters interact]”) . b. Apply grade 5 Reading standards to informational texts (e.g., “explain how an author uses reasons and evidence to support particular points in a text, identifying which reasons and evidence support which point[s]”).	3
WL.5.RW	Writing – Range of Writing	
ELA.5.29	Write routinely over extended time frames (time for research, reflection, and revision) and shorter time frames (a single sitting or a day or two) for a range of discipline-specific tasks, purposes, and audiences.	4
WL.5.CSE	Language - Conventions of Standard English	
ELA.5.36	Demonstrate command of the conventions of Standard English grammar and usage when writing or speaking. a. Explain the function of conjunctions, prepositions, and interjections in general and their function in particular sentences. b. Form and use the perfect (e.g., I had walked; I have walked; I will have walked) verb tenses. c. Use verb tense to convey various times, sequences, states, and conditions. d. Recognize and correct inappropriate shifts in verb tense. e. Use correlative conjunctions (e.g., either/or, neither/nor).	2
ELA.5.37	a. Demonstrate command of the conventions of Standard English capitalization, punctuation, and spelling when writing. a. Use punctuation to separate items in a series. b. Use a comma to separate an introductory element from the rest of the sentence. c. Use a comma to set off the words yes and no (e.g., Yes, thank you), to set off a tag question from the rest of the sentence (e.g., It’s true, isn’t it?), and to indicate direct address (e.g., Is that you, Steve?). d. Use underlining, quotation marks, or italics to indicate titles of works. e. Spell grade-appropriate words correctly, consulting references as needed.	1

WL.5.KL	Language - Knowledge of Language	
ELA.5.38	Use knowledge of language and its conventions when writing, speaking, reading, or listening. a. Expand, combine, and reduce sentences for meaning, reader/listener interest, and style. b. Compare and contrast the varieties of English (e.g., dialects, registers) used in stories, dramas, or poems.	2
WL.5.VAU	Language - Vocabulary Acquisition and Use	
ELA.5.39	Determine or clarify the meaning of unknown and multiple-meaning words and phrases based on grade 5 reading and content, choosing flexibly from a range of strategies. a. Use context (e.g., cause/effect relationships and comparisons in text) as a clue to the meaning of a word or phrase. b. Use common, grade-appropriate Greek and Latin affixes and roots as clues to the meaning of a word (e.g., photograph, photosynthesis). b. Consult reference materials (e.g., dictionaries, glossaries, thesauruses), both print and digital, to find the pronunciation and determine or clarify the precise meaning of key words and phrases.	2
ELA.5.40	Demonstrate understanding of figurative language, word relationships, and nuances in word meanings. a. Interpret figurative language, including similes and metaphors, in context. b. Recognize and explain the meaning of common idioms, adages, and proverbs. c. Use the relationship between particular words (e.g., synonyms, antonyms, homographs) to better understand each of the words.	3
ELA.5.41	Acquire and use accurately grade-appropriate general academic and domain-specific words and phrases, including those that signal contrast, addition, and other logical relationships (e.g., however, although, nevertheless, similarly, moreover, in addition).	2

	College and Career Readiness Standards for Grade 6 ELA	
Standard	Description	DOK
LR.6	Grade 6 Literary Reading	
LR.6.KID	Key Ideas and Details	
ELA.6.1	Cite textual evidence to support analysis of what the literary text says explicitly as well as inferences drawn from the text.	2
ELA.6.2	Determine a theme or central idea of a literary text and how it is conveyed through particular details; provide a summary of the text distinct from personal opinions or judgments.	2
ELA.6.3	Describe how a particular story's or drama's plot unfolds in a series of episodes as well as how the characters respond or change as the plot moves toward a resolution.	2
LR.6.CS	Craft and Structure	
ELA.6.7	Determine the meaning of words and phrases as they are used in a literary text, including figurative and connotative meanings; analyze the impact of a specific word choice on meaning and tone.	2
ELA.6.8	Analyze how a particular sentence, chapter, scene, or stanza fits into the overall structure of a literary text and contributes to the development of the theme, setting, or plot.	2
ELA.6.9	Explain how an author develops the point of view of the narrator or speaker in a literary text.	2
LR.6.IKI	Integration of Knowledge and Ideas	
ELA.6.13	Compare and contrast the experience of reading a story, drama, or poem to listening to or viewing an audio, video, or live version of the literary text, including contrasting what is “seen” and “heard” when reading the text to what they perceived when listening or watching.	3
ELA.6.14	Compare and contrast literary texts in different forms or genres (e.g., stories and poems; historical novels and fantasy stories) in terms of their approaches to similar themes and topics.	3

IR.6	Grade 6 Informational Reading	
IR.6.KID	Key Ideas and Details	
ELA.6.4	Cite textual evidence to support analysis of what the informational text says explicitly as well as inferences drawn from the text.	2
ELA.6.5	Determine a central idea of an informational text and how it is conveyed through particular details; provide a summary of the text distinct from personal opinions or judgments.	2
ELA.6.6	Analyze in detail how a key individual, event, or idea is introduced, illustrated, and developed in an informational text (e.g., through examples or anecdotes).	2
IR.6.CS	Craft and Structure.	
ELA.6.10	Determine the meaning of words and phrases as they are used in an informational text, including figurative, connotative, and technical meanings.	2
ELA.6.11	Analyze how a particular sentence, paragraph, chapter, or section fits into the overall structure of an informational text and contributes to the development of the ideas.	2
ELA.6.12	Determine an author's point of view or purpose in an informational text and explain how it is communicated in the text.	2
IR.6.IKI	Integration of Knowledge and Ideas	
ELA.6.15	Integrate information presented in different media or formats (e.g., visually and/or quantitatively) and in words to develop a coherent understanding of a topic or issue.	3
ELA.6.16	Trace and evaluate the argument and specific claims in an informational text, distinguishing claims that are supported by reasons and evidence from claims that are not.	3
ELA.6.17	Compare and contrast two author's presentation of events (e.g., a memoir written by and a biography on the same person) in informational text.	3

WL.6	Grade 6 Writing and Language	
WL.6.TTP	Writing - Text Types and Purposes	
ELA.6.20	Write arguments to support claims with clear reasons and relevant evidence. a. Introduce claim(s) and organize the reasons and evidence clearly. b. Support claim(s) with clear reasons and relevant evidence, using credible sources and demonstrating an understanding of the topic or text. c. Use words, phrases, and clauses to clarify the relationships among claim(s) and reasons. d. Establish and maintain a formal style. e. Provide a concluding statement or section that follows from the argument presented.	3
ELA.6.21	Write informative/explanatory texts to examine a topic and convey ideas, concepts, and information through the selection, organization, and analysis of relevant content. a. Introduce a topic; organize ideas, concepts, and information using strategies such as definition, classification, comparison/contrast, and cause/effect; include formatting (e.g., headings), graphics (e.g., charts, tables), and multimedia when useful to aiding comprehension. b. Develop the topic with relevant facts, definitions, concrete details, quotations, or other information and examples. c. Use appropriate transitions to clarify the relationships among ideas and concepts. d. Use precise language and domain-specific vocabulary to inform about or explain the topic. e. Establish and maintain a formal style. f. Provide a concluding statement or section that follows from the information or explanation presented.	3
ELA.6.22	Write narratives to develop real or imagined experiences or events using effective technique, relevant descriptive details, and well-structured event sequences. a. Engage and orient the reader by establishing a context and introducing a narrator and/or characters; organize an event sequence that unfolds naturally and logically. b. Use narrative techniques, such as dialogue, pacing, and description, to develop experiences, events, and/or characters. c. Use a variety of transition words, phrases, and clauses to convey sequence and signal shifts from one time frame or setting to another. d. Use precise words and phrases, relevant descriptive details, and sensory language to convey experiences and events. e. Provide a conclusion that follows from the narrated experiences or events.	3

WL.6.PDW	Writing - Production and Distribution of Writing	
ELA.6.23	Produce clear and coherent writing in which the development, organization, and style are appropriate to task, purpose, and audience. (Grade-specific expectations for writing types are defined in Text Types and Purposes.)	3
ELA.6.24	With some guidance and support from peers and adults, develop and strengthen writing as needed by planning, revising, editing, rewriting, or trying a new approach. (Editing for conventions should demonstrate command of Language standards up to and including grade 6.)	3
ELA.6.25	Use technology, including the Internet, to produce and publish writing as well as to interact and collaborate with others; demonstrate sufficient command of keyboarding skills to type efficiently and accurately.	2
WL.6.RBPK	Writing - Research to Build and Present Knowledge	
ELA.6.26	Conduct short research projects to answer a question, drawing on several sources and refocusing the inquiry when appropriate.	3
ELA.6.27	Gather relevant information from multiple print and digital sources; assess the credibility of each source; quote or paraphrase the data and conclusions of others while avoiding plagiarism and providing basic bibliographic information for sources.	3
ELA.6.28	Draw evidence from literary or informational texts to support analysis, reflection, and research. a. Apply grade 6 Reading standards to literature (e.g., "Compare and contrast texts in different forms or genres [e.g., stories and poems; historical novels and fantasy stories] in terms of their approaches to similar themes and topics"). b. Apply grade 6 Reading standards to literary nonfiction (e.g., "Trace and evaluate the argument and specific claims in a text, distinguishing claims that are supported by reasons and evidence from claims that are not").	3
WL.6.RW	Writing - Range of Writing	
ELA.6.29	Write routinely over extended time frames (time for research, reflection, and revision) and shorter time frames (a single sitting or a day or two) for a range of discipline-specific tasks, purposes, and audiences.	4
WL.6.CSE	Language - Conventions of Standard English	
ELA.6.36	Demonstrate command of the conventions of Standard English grammar and usage when writing or speaking. a. Ensure that pronouns are in the proper case (subjective, objective, and possessive). b. Use intensive pronouns (e.g., myself, ourselves). c. Recognize and correct inappropriate shifts in pronoun number and person. d. Recognize and correct vague pronouns (i.e., ones with unclear or ambiguous antecedents). e. Recognize variations from Standard English in their own and others' writing and speaking, and identify and use strategies to improve expression in conventional language.	1
ELA.6.37	Demonstrate command of the conventions of Standard English capitalization, punctuation, and spelling when writing. a. Use punctuation (commas, parentheses, dashes) to set off nonrestrictive/parenthetical elements. b. Spell correctly.	1

WL.6.KL	Language - Knowledge of Language	
ELA.6.38	Use knowledge of language and its conventions when writing, speaking, reading, or listening. a. Vary sentence patterns for meaning, reader/listener interest, and style. b. Maintain consistency in style and tone.	2
WL.6.VAU	Language - Vocabulary Acquisition and Use	
ELA.6.39	Determine or clarify the meaning of unknown and multiple-meaning words and phrases based on grade 6 reading and content, choosing flexibly from a range of strategies. a. Use context (e.g., the overall meaning of a sentence or paragraph; a word's position or function in a sentence) as a clue to the meaning of a word or phrase. b. Use common, grade-appropriate Greek or Latin affixes and roots as clues to the meaning of a word (e.g., audience, auditory, audible). c. Consult reference materials (e.g., dictionaries, glossaries, and/or thesauruses), both print and digital, to find the pronunciation of a word or determine or clarify its precise meaning or its part of speech. d. Verify the preliminary determination of the meaning of a word or phrase (e.g., by checking the inferred meaning in context or in a dictionary).	2
ELA.6.40	Demonstrate understanding of figurative language, word relationships, and nuances in word meanings. a. Interpret figures of speech (e.g., personification) in context. b. Use the relationship between particular words (e.g., cause/effect, part/whole, or item/category) to better understand each of the words. c. Distinguish among the connotations (associations) of words with similar denotations (definitions) (e.g., stingy, scrimping, economical, frugal, and thrifty).	2
ELA.6.41	Acquire and use accurately grade-appropriate general academic and domain-specific words and phrases; gather vocabulary knowledge when considering a word or phrase important to comprehension or expression.	2

	West Virginia College and Career Readiness Standards for Grade 7 ELA	
Standard	Description	DOK
LR.7	Grade 7 Literary Reading	
LR.7.KID	Key Ideas and Details	
ELA.7.1	Cite several pieces of textual evidence to support analysis of what the literary text says explicitly as well as inferences drawn from the text.	2
ELA.7.2	Determine a theme or central idea of a literary text and analyze its development over the course of the text; provide an objective summary of the text.	3
ELA.7.3	Analyze how particular elements of a story or drama interact (e.g., how setting shapes the characters or plot).	3
LR.7.CS	Craft and Structure	
ELA.7.7	Determine the meaning of words and phrases as they are used in a literary text, including figurative and connotative meanings; analyze the impact of rhymes and other repetitions of sounds (e.g., alliteration) on a specific verse or stanza of a poem or section of a story or drama.	3
ELA.7.8	Analyze how a drama's or poem's form or structure (e.g. soliloquy, sonnet), contributes to its meaning.	3
ELA.7.9	Analyze how an author develops and contrasts the points of view of different characters or narrators in a literary text.	3
LR.7.CS	Integration of Knowledge and Ideas	
ELA.7.13	Compare and contrast a written story, drama, or poem to its audio, filmed, staged, or multimedia version, analyzing the effects of techniques unique to each medium (e.g., lighting, sound, color, or camera focus and angles in a film).	3
ELA.7.14	Compare and contrast a fictional portrayal of a time, place, or character and a historical account of the same period as a means of understanding how authors of fiction use or alter history.	3

IR.7	Grade 7 Informational Reading	
IR.7.KID	Key Ideas and Details	
ELA.7.4	Cite several pieces of textual evidence to support analysis of what the informational text says explicitly as well as inferences drawn from the text.	2
ELA.7.5	Determine two or more central ideas in an informational text and analyze their development over the course of the text; provide an objective summary of the text.	3
ELA.7.6	Analyze the interactions between individuals, events, and ideas in an informational text (e.g., how ideas influence individuals or events, or how individuals influence ideas or events).	3
IR.7.CS	Craft and Structure.	
ELA.7.10	Determine the meaning of words and phrases as they are used in an informational text, including figurative, connotative, and technical meanings; analyze the impact of a specific word choice on meaning and tone.	3
ELA.7.11	Analyze the structure an author uses to organize an informational text, including how the major sections contribute to the whole and to the development of the ideas.	3
ELA.7.12	Determine an author's point of view or purpose in an informational text and analyze how the author distinguishes his or her position from that of others.	3
IR.7.IKI	Integration of Knowledge and Ideas	
ELA.7.15	Compare and contrast a text to an audio, video, or multimedia version of the informational text, analyzing each medium's portrayal of the subject (e.g., how the delivery of a speech affects the impact of the words).	3
ELA.7.16	Trace and evaluate the argument and specific claims in an informational text, assessing whether the reasoning is sound and the evidence is relevant and sufficient to support the claims.	3
ELA.7.17	Analyze how two or more authors writing about the same topic shape their presentations of key information by emphasizing different evidence or advancing different interpretations of facts.	3

WL.7	Grade 7 Writing and Language	
WL.7.TTP	Writing - Text Types and Purposes	
ELA.7.20	Write arguments to support claims with clear reasons and relevant evidence. a. Introduce claim(s), acknowledge alternate or opposing claims, and organize the reasons and evidence logically. b. Support claim(s) with logical reasoning and relevant evidence, using accurate, credible sources and demonstrating an understanding of the topic or text. c. Use words, phrases, and clauses to create cohesion and clarify the relationships among claim(s), reasons, and evidence. d. Establish and maintain a formal style. e. Provide a concluding statement or section that follows from and supports the argument presented.	3
ELA.7.21	Write informative/explanatory texts to examine a topic and convey ideas, concepts, and information through the selection, organization, and analysis of relevant content. a. Introduce a topic clearly, previewing what is to follow; organize ideas, concepts, and information, using strategies such as definition, classification, comparison/contrast, and cause/effect; include formatting (e.g., headings), graphics (e.g., charts, tables), and multimedia when useful to aid comprehension. b. Develop the topic with relevant facts, definitions, concrete details, quotations, or other information and examples. c. Use appropriate transitions to create cohesion and clarify the relationships among ideas and concepts. d. Use precise language and domain-specific vocabulary to inform about or explain the topic. e. Establish and maintain a formal style. f. Provide a concluding statement or section that follows from and supports the information or explanation presented.	3
ELA.7.22	Write narratives to develop real or imagined experiences or events using effective technique, relevant descriptive details, and well-structured event sequences. a. Engage and orient the reader by establishing a context and point of view and introducing a narrator and/or characters; organize an event sequence that unfolds naturally and logically. b. Use narrative techniques, such as dialogue, pacing, and description, to develop experiences, events, and/or characters. c. Use a variety of transition words, phrases, and clauses to convey sequence and signal shifts from one time frame or setting to another. d. Use precise words and phrases, relevant descriptive details, and sensory language to capture the action and convey experiences and events. e. Provide a conclusion that follows from and reflects on the narrated experiences or events.	3

WL.7.PDW	Writing - Production and Distribution of Writing	
ELA.7.23	Produce clear and coherent writing in which the development, organization, and style are appropriate to task, purpose, and audience. (Grade-specific expectations for writing types are defined in Text Types and Purposes.)	3
ELA.7.24	With some guidance and support from peers and adults, develop and strengthen writing as needed by planning, revising, editing, rewriting, or trying a new approach, focusing on how well purpose and audience have been addressed. (Editing for conventions should demonstrate command of Language standards up to and including grade 7.)	3
ELA.7.25	Use technology, including the Internet, to produce and publish writing, link to and cite sources, and interact and collaborate with others.	2
WL.7.RBPK	Writing - Research to Build and Present Knowledge	
ELA.7.26	Conduct short research projects to answer a question, drawing on several sources and generating additional related, focused questions for further research and investigation.	3
ELA.7.27	Gather relevant information from multiple print and digital sources, using search terms effectively; assess the credibility and accuracy of each source; and quote or paraphrase the data and conclusions of others while avoiding plagiarism and following a standard format for citation (e.g., MLA or APA).	3
ELA.7.28	Draw evidence from literary or informational texts to support analysis, reflection, and research. a. Apply <i>grade 7 Reading standards</i> to literature (e.g., "Compare and contrast a fictional portrayal of a time, place, or character and a historical account of the same period as a means of understanding how authors of fiction use or alter history"). b. Apply <i>grade 7 Reading standards</i> to nonfiction and other informational texts (e.g. "Trace and evaluate the argument and specific claims in a text, assessing whether the reasoning is sound and the evidence is relevant and sufficient to support the claims").	3
WL.7.RW	Writing - Range of Writing	
ELA.7.29	Write routinely over extended time frames (time for research, reflection, and revision) and shorter time frames (a single sitting or a day or two) for a range of discipline-specific tasks, purposes, and audiences.	4

WL.7.CSE	Language - Conventions of Standard English	
ELA.7.36	Demonstrate command of the conventions of standard English grammar and usage when writing or speaking. a. Explain the function of phrases and clauses in general and their function in specific sentences. b. Choose among simple, compound, complex, and compound-complex sentences to signal differing relationships among ideas. c. Place phrases and clauses within a sentence, recognizing and correcting misplaced and dangling modifiers.	2
ELA.7.37	Demonstrate command of the conventions of standard English capitalization, punctuation, and spelling when writing. a. Use a comma to separate coordinate adjectives (e.g., It was a fascinating, enjoyable movie). b. Spell correctly.	1
WL.7.KL	Language - Knowledge of Language	
ELA.7.38	Use knowledge of language and its conventions when writing, speaking, reading, or listening. a. Choose language that expresses ideas precisely and concisely, recognizing and eliminating wordiness and redundancy.	2
WL.7.VAU	Language - Vocabulary Acquisition and Use	
ELA.7.39	Determine or clarify the meaning of unknown and multiple-meaning words and phrases based on grade 7 reading and content, choosing flexibly from a range of strategies. a. Use context (e.g., the overall meaning of a sentence or paragraph; a word's position or function in a sentence) as a clue to the meaning of a word or phrase. b. Use common, grade-appropriate Greek or Latin affixes and roots as clues to the meaning of a word (e.g., <i>belligerent</i> , <i>bellicose</i> , <i>rebel</i>). c. Consult general and specialized reference materials (e.g., dictionaries, glossaries, thesauruses), both print and digital, to find the pronunciation of a word or determine or clarify its precise meaning or its part of speech. d. Verify the preliminary determination of the meaning of a word or phrase (e.g., by checking the inferred meaning in context or in a dictionary).	2
ELA.7.40	Demonstrate understanding of figurative language, word relationships, and nuances in word meanings. a. Interpret figures of speech (e.g., literary or mythological allusions) in context. b. Use the relationship between particular words (e.g., synonym/antonym or analogy) to better understand each of the words. c. Distinguish among the connotations (associations) of words with similar denotations (definitions) (e.g., refined, respectful, polite, diplomatic, condescending).	2
ELA.7.41	Acquire and use accurately grade-appropriate general academic and domain-specific words and phrases; gather vocabulary knowledge when considering a word or phrase important to comprehension or expression.	2

	West Virginia College and Career Readiness Standards for Grade 8 ELA	
Standard	Description	DOK
LR.8	Grade 8 Literary Reading	
LR.8.KID	Key Ideas and Details	
ELA.8.1	Cite the textual evidence that most strongly supports an analysis of what the literary text says explicitly as well as inferences drawn from the text.	2
ELA.8.2	Determine a theme or central idea of a literary text and analyze its development over the course of the text, including its relationship to the characters, setting, and plot; provide an objective summary of the text.	3
ELA.8.3	Analyze how particular lines of dialogue or incidents in a story or drama propel the action, reveal aspects of a character, or provoke a decision.	3
LR.8.CS	Craft and Structure	
ELA.8.7	Determine the meaning of words and phrases as they are used in a literary text, including figurative and connotative meanings; analyze the impact of specific word choices on meaning and tone, including analogies or allusions to other texts.	3
ELA.8.8	Compare and contrast the structure of two or more literary texts and analyze how the differing structure of each text contributes to its meaning and style.	3
ELA.8.9	Analyze how differences in the points of view of the characters and the audience or reader (e.g., created through the use of dramatic irony) create such effects as suspense or humor in a literary text.	3
LR.8.IKI	Integration of Knowledge and Ideas	
ELA.8.13	Analyze the extent to which a filmed or live production of a story or drama stays faithful to or departs from the text or script, evaluating the choices made by the director or actors.	3
ELA.8.14	Analyze how a modern work of fiction draws on themes, patterns of events, or character types from myths, traditional stories or religious works such as the Bible, including describing how the material is transformed in the modern work (e.g., how a modern interpretation of a Shakespearean text draws from the original text).	3

IR.8	Grade 8 Informational Reading	
IR.8.KID	Key Ideas and Details	
ELA.8.4	Cite the textual evidence that most strongly supports an analysis of what the informational text says explicitly as well as inferences drawn from the text.	2
ELA.8.5	Determine a central idea of an informational text and analyze its development over the course of the text, including its relationship to supporting ideas; provide an objective summary of the text.	3
ELA.8.6	Analyze how an informational text makes connections among and distinctions between individuals, ideas, or events (e.g., through comparisons, analogies, or categories).	3
IR.8.CS	Craft and Structure	
ELA.8.10	Determine the meaning of words and phrases as they are used in an informational text, including figurative, connotative, and technical meanings; analyze the impact of specific word choices on meaning and tone, including analogies or allusions to other texts.	3
ELA.8.11	Analyze in detail the structure of a specific paragraph in an informational text, including the role of particular sentences in developing and refining a key concept.	3
ELA.8.12	Determine an author's point of view or purpose in an informational text and analyze how the author acknowledges and responds to conflicting evidence or viewpoints.	3
IR.8.IKI	Integration of Knowledge and Ideas	
ELA.8.15	Evaluate the advantages and disadvantages of using different mediums (e.g., print or digital text, video, and/or multimedia) to present a particular topic or idea.	3
ELA.8.16	Delineate and evaluate the argument and specific claims in an informational text, assessing whether the reasoning is sound and the evidence is relevant and sufficient; recognize when irrelevant evidence is introduced.	3
ELA.8.17	Analyze a case in which two or more informational texts provide conflicting information on the same topic and identify where the texts disagree on matters of fact or interpretation.	2

WL.8	Grade 8 Writing and Language	
WL.8.TTP	Writing - Text Types and Purposes	
ELA.8.20	Write arguments to support claims with clear reasons and relevant evidence. a. Introduce claim(s), acknowledge and distinguish the claim(s) from alternate or opposing claims, and organize the reasons and evidence logically. b. Support claim(s) with logical reasoning and relevant evidence, using accurate, credible sources and demonstrating an understanding of the topic or text. c. Use words, phrases, and clauses to create cohesion and clarify the relationships among claim(s), counterclaims, reasons, and evidence. d. Establish and maintain a formal style. e. Provide a concluding statement or section that follows from and supports the argument presented.	3
ELA.8.21	Write informative/explanatory texts to examine a topic and convey ideas, concepts, and information through the selection, organization, and analysis of relevant content. a. Introduce a topic clearly, previewing what is to follow; organize ideas, concepts, and information into broader categories; include formatting (e.g., headings), graphics (e.g., charts, tables), and multimedia when useful to aiding comprehension. b. Develop the topic with well-chosen, relevant facts, definitions, concrete details, quotations, or other information and examples. c. Use appropriate and varied transitions to create cohesion and clarify the relationships among ideas and concepts. d. Use precise language and domain-specific vocabulary to inform about or explain the topic. e. Establish and maintain a formal style. f. Provide a concluding statement or section that follows from and supports the information or explanation presented.	3
ELA.8.22	Write narratives to develop real or imagined experiences or events using effective technique, relevant descriptive details, and well-structured event sequences. a. Engage and orient the reader by establishing a context and point of view and introducing a narrator and/or characters; organize an event sequence that unfolds naturally and logically. b. Use narrative techniques, such as dialogue, pacing, description, and reflection, to develop experiences, events, and/or characters. c. Use a variety of transition words, phrases, and clauses to convey sequence, signal shifts from one time frame or setting to another, and show the relationships among experiences and events. d. Use precise words and phrases, relevant descriptive details, and sensory language to capture the action and convey experiences and events. e. Provide a conclusion that follows from and reflects on the narrated experiences or events.	3

WL.8.PDW	Writing - Production and Distribution of Writing	
ELA.8.23	Produce clear and coherent writing in which the development, organization, and style are appropriate to task, purpose, and audience. (Grade-specific expectations for writing types are defined in Text Types and Purposes.)	3
ELA.8.24	With some guidance and support from peers and adults, develop and strengthen writing as needed by planning, revising, editing, rewriting, or trying a new approach, focusing on how well purpose and audience have been addressed. (Editing for conventions should demonstrate command of Language standards up to and including grade 8.)	3
ELA.8.25	Use technology, including the Internet, to produce and publish writing, present the relationships between information and ideas efficiently, and interact and collaborate with others.	2
WL.8.RBPK	Writing - Research to Build and Present Knowledge	
ELA.8.26	Conduct short research projects to answer a question (including a self-generated question), drawing on several sources and generating additional related, focused questions that allow for multiple avenues of exploration.	3
ELA.8.27	Gather relevant information from multiple print and digital sources, using search terms effectively; assess the credibility and accuracy of each source; and quote or paraphrase the data and conclusions of others while avoiding plagiarism and following a standard format for citation (e.g., MLA or APA).	3
ELA.8.28	Draw evidence from literary or informational texts to support analysis, reflection, and research. a. Apply grade 8 Reading standards to literature (e.g., “Analyze how a modern work of fiction draws on themes, patterns of events, or character types from myths, traditional stories, or religious works, such as the Bible, including describing how the material is transformed in the modern work (e.g., how a modern interpretation of a Shakespearean text draws from the original text)”). b. Apply grade 8 Reading standards to literary nonfiction (e.g., “Delineate and evaluate the argument and specific claims in a text, assessing whether the reasoning is sound and the evidence is relevant and sufficient; recognize when irrelevant evidence is introduced.”).	3
WL.8.RW	Writing - Range of Writing	
ELA.8.29	Write routinely over extended time frames (time for research, reflection, and revision) and shorter time frames (a single sitting or a day or two) for a range of discipline-specific tasks, purposes, and audiences.	4
WL.8.CSE	Language - Conventions of Standard English	
ELA.8.36	Demonstrate command of the conventions of Standard English grammar and usage when writing or speaking. a. Explain the function of verbals (gerunds, participles, infinitives) in general and their function in particular sentences.	2

ELA.8.37	Demonstrate command of the conventions of Standard English capitalization, punctuation, and spelling when writing. a. Use punctuation (comma, ellipsis, or dash) to indicate a pause or break. b. Use an ellipsis to indicate an omission. c. Spell correctly.	1
WL.8.KL	Language - Knowledge of Language	
ELA.8.38	Use knowledge of language and its conventions when writing, speaking, reading, or listening. a. Use verbs in the active and passive voice (e.g., emphasizing the actor or the action). b. Use verbs in the indicative, imperative, interrogative, conditional and subjunctive mood to achieve particular effects (e.g., expressing uncertainty or describing a state contrary to fact). c. Recognize and correct inappropriate shifts in verb voice and mood.	2
WL.8.VAU	Language - Vocabulary Acquisition and Use	
ELA.8.39	Determine or clarify the meaning of unknown and multiple-meaning words or phrases based on grade 8 reading and content, choosing flexibly from a range of strategies. a. Use context (e.g., the overall meaning of a sentence or paragraph or a word's position or function in a sentence) as a clue to the meaning of a word or phrase. b. Use common, grade-appropriate Greek or Latin affixes and roots as clues to the meaning of a word (e.g., precede, recede, and secede). c. Consult general and specialized reference materials (e.g., dictionaries, glossaries, thesauruses), both print and digital, to find the pronunciation of a word or determine or clarify its precise meaning or its part of speech. d. Verify the initial determination of the meaning of a word or phrase (e.g. by checking the inferred meaning in context in a dictionary).	2
ELA.8.40	Demonstrate understanding of figurative language, word relationships, and nuances in word meanings. a. Interpret figures of speech (e.g. verbal irony, puns) in context. b. Use the relationship between particular words to better understand each of the words. c. Distinguish among the connotations (associations) of words with similar denotations (definitions) (e.g., bullheaded, willful, firm, persistent, resolute).	2
ELA.8.41	Acquire and use accurately grade-appropriate general academic and domain-specific words and phrases; gather vocabulary knowledge when considering a word or phrase important to comprehension or expression.	2

Appendix B

Data Analysis Tables

West Virginia College- and Career- Readiness Standards for English Language Arts Grades 3-8 and Sample Test Events

2019

Brief Explanation of Data in the Alignment Tables by Column

Tables x.1

Domain # Number of Domains for each Reporting Category (RC)

Standards # Average number of standards for reviewers. If the number is greater than the actual number in the standard, then at least one reviewer coded an item for the Domain or RC but did not find any standard in the Domain that corresponded to the item.

Level The Depth-of-Knowledge levels coded by the reviewers for the standards for the RC.

Num of Stds by Level The number of standards coded at each DOK level
% w/in RC by Level The percent of standards within the RC coded at each DOK level

Hits

Mean & SD Mean and standard deviation number of items reviewers coded as corresponding to reporting category. The total is the total number of coded hits.

Categorical Concurrence:

“Yes” indicates that the reporting category met the acceptable level for criterion.

“Yes” if mean is six or more.

“Weak” if mean is five to six.

“No” if mean is less than five.

Tables x.2

First five columns repeat columns from Table 1.

DOK Level of Item Gives percentages of DOK levels of items in relation to standards

Mean percent and standard deviation of items coded as “under” the Depth-of-Knowledge level of the corresponding standard, as “at” (the same) the Depth-of-Knowledge level of the corresponding standard, and as “above” the Depth-of-Knowledge level of the corresponding standard.

DOK Consistency:

“Yes” indicates that 50% or more of the items were rated as “at” or “above” the Depth-of-Knowledge level of the corresponding standards.

“Weak” indicates that 40% to 50% of the items were rated as “at” or “above” the Depth-of-Knowledge level of the corresponding standards.

“No” indicates that less than 40% items were rated as “at” or “above” the Depth-of-Knowledge level of the corresponding standards.

Tables x.3

First five columns repeat columns from Table 1 and 2.

Range of Standards

Num Stds Hit Average number and standard deviation of the standards hit coded by reviewers.

% of Total Average percent and standard deviation of the total standards that had at least one item coded.

Range of Know(ledge):

“Yes” indicates that 50% or more of the standards had at least one coded standard.

“Weak” indicates that 40% to 50% of the standards had at least one coded standard.

“No” indicates that 40% or less of the standards had at least one coded standard.

Balance Index:

Gives % Hits in Std/Ttl Hit Average and standard deviation of the percent of the items hit for a reporting category of total number of hits (see total under the Hits column).

Index Average and standard deviation of the Balance Index.

Note: $BALANCE\ INDEX = 1 - (\sum_{k=1} |1/(O) - I_{(k)} / (H)|) / 2$

Where O = Total number of standards hit for the reporting category

$I_{(k)}$ = Number of items hit corresponding to

standard (k)

H = Total number of items hit for the reporting category

Balance of Representation:

“Yes” indicates that the Balance Index was .7 or above (items evenly distributed among standards).

“Weak” indicates that the Balance Index was .6 to .7 (a high percentage of items coded as corresponding to two or three standards).

“No” indicates that the Balance Index was .6 or less (a high percentage of items coded as corresponding to one standard.)

Tables x.4

Summary of if reporting category met the acceptable level for the four criteria by each standard.

Tables x.5

The DOK value for each assessment item given by each reviewer. The intraclass correlation for the group of reviewers is given on the last row.

Tables x.6

The DOK level and standard code assigned by each reviewer for each item.

Tables x.7

This lists for each standard all of the items coded by the group of reviewers as corresponding to the standard. The number of reviewers who coded the item is given in parentheses.

Tables x.8

This list for each item all of the standards coded by the group of reviewers as corresponding to the item. The number of reviewers who coded the standard is given in after the colon.

Tables x.9

This table can be used to compare approximately the DOK level of a standard to the average DOK level of the items reviewers assigned to the standard. This table is helpful to identify items with a lower DOK level that should be replaced by an item with a higher DOK level to improve the Depth-of-Knowledge Consistency. The DOK listed in the table for each item is generally the mode DOK for that item.

Grade 3 Batch 1 ELA West Virginia

Table 3.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 3 B1 Number of Assessment Items - 43*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.3 Grade 3 Literary Reading	3	8	2 3	5 3	62.5 37.5	23.33	0.82	YES
IR.3 Grade 3 Informational Reading	3	9	2 3	5 4	55.56 44.44	9.17	0.41	YES
WL.3 Grade 3 Writing and Language	7	15	1 2 3 4	2 5 7 1	13.33 33.33 46.67 6.67	25.67	1.37	YES
Total	13	32	1 2 3 4	2 15 14 1	6 47 44 3	58.17	1.17	

Table 3.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 3 B1 Number of Assessment Items - 43*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.3 Grade 3 Literary Reading	3	8	23.33	0.82	32.5	14	63.15	13	4.35	7	YES
IR.3 Grade 3 Informational Reading	3	9	9.17	0.41	35	24	65	24	0	0	YES
WL.3 Grade 3 Writing and Language	7	15	25.67	1.37	0	0	84.53	6	15.47	6	YES
Total	13	32	58.17	1.17	18.62	9.7	72.78	9.7	8.6	5.3	

*Table 3.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers
WV ELA 2019 Grade 3 B1 Number of Assessment Items - 43*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
RL.3 Grade 3 Reading: Literature	3	8	23.33	0.82	6.5	0.55	81.25	6.85	YES	50	2	0.76	0.07	YES
RI.3 Grade 3 Reading: Informational Text	3	9	9.17	0.41	6	0	66.67	0	YES	20	1	0.84	0.01	YES
W.L.3 Grade 3 Writing and Language	7	15	25.67	1.37	5.17	0.41	34.44	2.72	NO	30	1	0.69	0.05	WEAK
Total	13	32	58.17	1.17	5.9	0.67	60.79	24		33	15	0.76	0.07	

*Table 3.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV ELA 2019 Grade 3 B1
Number of Assessment Items - 43*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
RL.3 Grade 3 Reading: Literature	YES	YES	YES	YES
RI.3 Grade 3 Reading: Informational Text	YES	YES	YES	YES
W.L.3 Grade 3 Writing and Language	YES	YES	NO	WEAK

Table 3.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 3 B1

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6	Reviewer 7
1	2		2	2	2	2	2
2	3		2	2	2	2	2
3	2		1	2	2	2	2
4	2		2	3	2	2	2
5	1		2	1	1	1	1
6	2		1	1	2	2	1
7	2		2	1	2	1	2
8	2		2	2	2	2	2
9	2		2	2	2	2	2
10	2		2	3	2	2	2
11	2		3	3	2	2	2
12	2		2	2	2	1	2
13	3		2	3	3	2	2
14	2		2	2	2	2	2
15	3		2	3	3	2	2
16	2		2	2	3	2	2
17	1		1	1	1	1	1
18	1		2	1	1	1	1
19	1		1	1	1	1	1
20	1		1	1	1	1	1
21	2		2	2	2	1	2
22	2		2	2	2	2	2
23	2		1	1	2	1	2
24	2		2	3	2	2	2
25	3		2	3	2	2	2
26	3		3	3	3	2	2
27	1		2	2	1	1	2
28	2		1	1	2	1	2
29	3		3	3	3	2	2
30	2		2	2	2	1	2
31	3		2	3	2	2	2

32	3		2	2	3	2	2
33	2		2	2	2	1	2
34	2		2	2	2	1	2
35	2		2	2	3	2	2
36	2		2	2	2	1	2
37	1		1	1	1	1	1
38	1		1	1	1	1	1
39	1		2	1	1	1	1
40	1		1	1	1	1	1
41	1		1	2	1	1	1
42	3		3	3	3	3	3
43	3		3	3	3	3	3

Intraclass correlation - .9423

Pairwise Comparison - 0.7

Table 3.6
DOK Levels and Objectives Code by Each Reviewer
WV ELA 2019 Grade 3 B1

Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.3.8			2	ELA.3.8			2	ELA.3.8			2	ELA.3.1			2	ELA.3.8			2	ELA.3.8		
2	2	ELA.3.7			2	ELA.3.7			3	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.3 9		
3	1	ELA.3.8			2	ELA.3.8			2	ELA.3.1 1			2	ELA.3.8			2	ELA.3.8			2	ELA.3.8		
4	2	ELA.3.3			2	ELA.3.9			2	ELA.3.9			2	ELA.3.1			2	ELA.3.9			3	ELA.3.9		
5	2	ELA.3.2			1	ELA.3.1			1	ELA.3.1			1	ELA.3.1			1	ELA.3.1			1	ELA.3.1		
6	1	ELA.3.1			2	ELA.3.1			2	ELA.3.1			1	ELA.3.1			2	ELA.3.1			1	ELA.3.7		
7	2	ELA.3.3			1	ELA.3.1			2	ELA.3.1			2	ELA.3.3			2	ELA.3.3			1	ELA.3.3		
8	2	ELA.3.2			2	ELA.3.2			2	ELA.3.2			2	ELA.3.2			2	ELA.3.2			2	ELA.3.2		
9	2	ELA.3.1			2	ELA.3.1			2	ELA.3.1			2	ELA.3.1			2	ELA.3.1			2	ELA.3.1		
10	2	ELA.3.1			2	ELA.3.8			2	ELA.3.8			2	ELA.3.3			2	ELA.3.8			3	ELA.3.3		
11	3	ELA.3.9			2	ELA.3.9			2	ELA.3.9			2	ELA.3.1	ELA.3.3		2	ELA.3.3			3	ELA.3.9	ELA.3.3	
12	2	ELA.3.9			1	ELA.3.3			2	ELA.3.3			2	ELA.3.1			2	ELA.3.1			2	ELA.3.3		
13	2	ELA.3.1 4			2	ELA.3.3			3	ELA.3.1 4			2	ELA.3.1 4			3	ELA.3.1 4			3	ELA.3.2		
14	2	ELA.3.7			2	ELA.3.7	ELA.3.3		2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7		
15	2	ELA.3.3			2	ELA.3.3			3	ELA.3.3			2	ELA.3.3			3	ELA.3.3			3	ELA.3.3		
16	2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			3	ELA.3.3			2	ELA.3.7		
17	1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6		
18	2	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6		
19	1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 7			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6		
20	1	ELA.3.3 7			1	ELA.3.3 7			1	ELA.3.3 7			1	ELA.3.3 7			1	ELA.3.3 7			1	ELA.3.3 7		
21	2	ELA.3.4			1	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4		
22	2	ELA.3.5			2	ELA.3.5			2	ELA.3.5			2	ELA.3.5			2	ELA.3.5			2	ELA.3.5		
23	1	ELA.3.1 0			1	ELA.3.1 0			2	ELA.3.1 0			2	ELA.3.1 0			2	ELA.3.1 0			1	ELA.3.1 0		

24	2	ELA.3.1 1		2	ELA.3.1 1		2	ELA.3.1 1		2	ELA.3.5		2	ELA.3.1 1		3	ELA.3.6		
25	2	ELA.3.1 5		2	ELA.3.1 5		3	ELA.3.1 5		2	ELA.3.1 5		2	ELA.3.1 5		3	ELA.3.1 5		
26	3	ELA.3.1 7		2	ELA.3.1 7		3	ELA.3.1 7		2	ELA.3.1 7		3	ELA.3.1 7		3	ELA.3.1 7		
27	2	ELA.3.1 1		1	ELA.3.4		1	ELA.3.4		2	ELA.3.6		1	ELA.3.4		2	ELA.3.4		
28	1	ELA.3.1 0		1	ELA.3.1 0		2	ELA.3.1 0		2	ELA.3.1 0		2	ELA.3.1 0		1	ELA.3.1 0		
29	3	ELA.3.1 7		2	ELA.3.1 7		3	ELA.3.1 7		2	ELA.3.1 7		3	ELA.3.1 7		3	ELA.3.1 7		
30	2	ELA.3.3		1	ELA.3.3		2	ELA.3.3		2	ELA.3.3		2	ELA.3.3		2	ELA.3.3		
31	2	ELA.3.1		2	ELA.3.3		3	ELA.3.3		2	ELA.3.1		2	ELA.3.2		3	ELA.3.2		
32	2	ELA.3.3		2	ELA.3.9		3	ELA.3.3		2	ELA.3.3		3	ELA.3.3		2	ELA.3.3		
33	2	ELA.3.7		1	ELA.3.7		2	ELA.3.7		2	ELA.3.7		2	ELA.3.7		2	ELA.3.7		
34	2	ELA.3.9		1	ELA.3.9		2	ELA.3.9		2	ELA.3.1		2	ELA.3.9		2	ELA.3.9		
35	2	ELA.3.3		2	ELA.3.3		2	ELA.3.3		2	ELA.3.3		3	ELA.3.3		2	ELA.3.1		
36	2	ELA.3.7		1	ELA.3.7		2	ELA.3.7		2	ELA.3.7		2	ELA.3.7		2	ELA.3.7		
37	1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		
38	1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		
39	2	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		
40	1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		
41	1	ELA.3.3 6	ELA.3.3 7	1	ELA.3.3 6	ELA.3.3 7	1	ELA.3.3 6	ELA.3.3 7	1	ELA.3.3 6	ELA.3.3 7	1	ELA.3.3 6	ELA.3.3 7	2	ELA.3.3 6	ELA.3.3 7	
42	3	ELA.3.2 1	ELA.3.3 8	3	ELA.3.3 8		3	ELA.3.2 1	ELA.3.3 8	3	ELA.3.2 1	ELA.3.3 8	3	ELA.3.2 1	ELA.3.3 8	3	ELA.3.2 1	ELA.3.3 8	
43	3	ELA.3.2 3	ELA.3.2 1	3	ELA.3.2 1	ELA.3.2 3	3	ELA.3.2 1	ELA.3.2 3	3	ELA.3.2 1	ELA.3.2 3	3	ELA.3.2 1	ELA.3.2 3	3	ELA.3.2 1	ELA.3.2 3	
Objective Pairwise Comparison: 0.75																			
Standard Pairwise Comparison: 0.98																			

Table 3.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV ELA 2019 Grade 3 B1

	Low			Medium				High					
	0			10.8				18					
LR.3													
LR.3.KID													
ELA.3.1	1(1)	4(1)	5(5)	6(5)	7(2)	9(6)	10(1)	11(1)	12(2)	31(2)	35(1)	34(1)	
ELA.3.2	31(2)	13(1)	8(6)	5(1)									
ELA.3.3	7(4)	4(1)	15(6)	16(1)	13(1)	12(3)	10(2)	31(3)	32(5)	30(6)	35(5)	14(1)	11(3)
LR.3.CS													
ELA.3.7	36(6)	33(6)	14(6)	16(5)	2(5)	6(1)							
ELA.3.8	1(5)	3(5)	10(3)										
ELA.3.9	11(4)	12(1)	4(4)	34(5)	32(1)								
LR.3.IKI													
ELA.3.13													
ELA.3.14	13(4)												
IR.3													
IR.3.KID													
ELA.3.4	21(6)	27(4)											
ELA.3.5	22(6)	24(1)											
ELA.3.6	24(1)	27(1)											
IR.3.CS													
ELA.3.10	28(6)	23(6)											
ELA.3.11	24(4)	3(1)	27(1)										
ELA.3.12													
IR.3.IKI													
ELA.3.15	25(6)												

ELA.3.16													
ELA.3.17	26(6)	29(6)											
WL.3													
WL.3.TTP													
ELA.3.20													
ELA.3.21	42(15)	43(18)											
ELA.3.22													
WL.3.PDW													
ELA.3.23	43(18)												
ELA.3.24													
ELA.3.25													
WL.3.RBPK													
ELA.3.26													
ELA.3.27													
WL.3.RW													
ELA.3.29													
WL.3.CSE													
ELA.3.36	39(6)	40(6)	41(12)	37(6)	17(6)	18(6)	19(10)						
ELA.3.37	19(2)	20(6)	38(12)	41(12)									
WL.3.KL													
ELA.3.38	42(18)												
WL.3.VAU													
ELA.3.39	2(1)												
ELA.3.40													
ELA.3.41													

Table 3.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 3 B1

Low 3.6		Medium 10.8		High 18	
1 10199-4978	ELA.3.1:1	ELA.3.8:5			
2 10199-4986	ELA.3.7:5	ELA.3.39:1			
3 10199-4987	ELA.3.8:5	ELA.3.11:1			
4 10199-4989	ELA.3.1:1	ELA.3.3:1	ELA.3.9:4		
5 10199-4992	ELA.3.1:5	ELA.3.2:1			
6 10199-4993	ELA.3.1:5	ELA.3.7:1			
7 10199-11976	ELA.3.1:2	ELA.3.3:4			
8 10199-11978	ELA.3.2:6				
9 10199-11979	ELA.3.1:6				
10 10199-11986	ELA.3.1:1	ELA.3.3:2	ELA.3.8:3		
11 10199-11987	ELA.3.1:1	ELA.3.3:3	ELA.3.9:4		
12 10199-11988	ELA.3.1:2	ELA.3.3:3	ELA.3.9:1		
13 10199-11989	ELA.3.2:1	ELA.3.3:1	ELA.3.14:4		
14 10199-11991	ELA.3.3:1	ELA.3.7:6			
15 10199-13975	ELA.3.3:6				
16 10199-13976	ELA.3.3:1	ELA.3.7:5			
17 10199-1173	ELA.3.36:6				
18 10199-1181	ELA.3.36:6				
19 10199-1185	ELA.3.36:10	ELA.3.37:2			
20 10199-1191	ELA.3.37:6				
21 REP10199-6897	ELA.3.4:6				
22 REP10199-6899	ELA.3.5:6				
23 REP10199-6907	ELA.3.10:6				
24 REP10199-6911	ELA.3.5:1	ELA.3.6:1	ELA.3.11:4		
25 10199-6913	ELA.3.15:6				
26 REP10199-6914	ELA.3.17:6				
27 REP10199-6902	ELA.3.4:4	ELA.3.6:1	ELA.3.11:1		
28 REP10199-6915	ELA.3.10:6				
29 REP10199-6909	ELA.3.17:6				
30 REP10199-6586	ELA.3.3:6				

31 REP10199-6591	ELA.3.1:2	ELA.3.2:2	ELA.3.3:3
32 REP10199-6592	ELA.3.3:5	ELA.3.9:1	
33 REP10199-6596	ELA.3.7:6		
34 REP10199-6598	ELA.3.1:1	ELA.3.9:5	
35 REP10199-6599	ELA.3.1:1	ELA.3.3:5	
36 REP10199-6601	ELA.3.7:6		
37 REP10199-1245	ELA.3.36:6		
38 REP10199-1246	ELA.3.37:12		
39 REP10199-1248	ELA.3.36:6		
40 REP10199-1250	ELA.3.36:6		
41 CONVENTIONS10199-14161	ELA.3.36:12	ELA.3.37:12	
42 ELABORATION10199-14161	ELA.3.21:15	ELA.3.38:18	
43 ORGANIZATION10199-14161	ELA.3.21:18	ELA.3.23:18	

Table 3.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 3 B1

Low DOK		Matched DOK		High DOK

LR.3													
LR.3.KID													
ELA.3.1: [2]	1:(1)[2]	4:(1)[2]	5:(5)[1]	6:(5)[2]	7:(2)[2]	9:(6)[2]	10:(1)[2]	11:(1)[2]	12:(2)[2]	31:(2)[2]	34:(1)[2]	35:(1)[2]	
ELA.3.2: [2]	5:(1)[2]	8:(6)[2]	13:(1)[3]	31:(2)[2]									
ELA.3.3: [3]	4:(1)[2]	7:(4)[2]	10:(2)[2]	11:(3)[2]	12:(3)[2]	13:(1)[2]	14:(1)[2]	15:(6)[2]	16:(1)[3]	30:(6)[2]	31:(3)[2]	32:(5)[2]	35:(5)[2]
LR.3.CS													
ELA.3.7: [2]	2:(5)[2]	6:(1)[1]	14:(6)[2]	16:(5)[2]	33:(6)[2]	36:(6)[2]							
ELA.3.8: [2]	1:(5)[2]	3:(5)[2]	10:(3)[2]										
ELA.3.9: [2]	4:(4)[2]	11:(4)[2]	12:(1)[2]	32:(1)[2]	34:(5)[2]								
LR.3.IKI													
ELA.3.13													
ELA.3.14: [3]	13:(4)[2]												
IR.3													
IR.3.KID													
ELA.3.4: [2]	21:(6)[2]	27:(4)[1]											
ELA.3.5: [2]	22:(6)[2]	24:(1)[2]											
ELA.3.6: [3]	24:(1)[3]	27:(1)[2]											
IR.3.CS													
ELA.3.10: [2]	23:(6)[2]	28:(6)[2]											
ELA.3.11: [2]	3:(1)[2]	24:(4)[2]	27:(1)[2]										
ELA.3.12													
IR.3.IKI													
ELA.3.15: [3]	25:(6)[2]												
ELA.3.16													
ELA.3.17: [3]	26:(6)[3]	29:(6)[3]											
WL.3													
WL.3.TTP													
ELA.3.20													

ELA.3.21: [3]	42:(15)[3]	43:(18)[3]											
ELA.3.22													
WL.3.PDW													
ELA.3.23: [3]	43:(18)[3]												
ELA.3.24													
ELA.3.25													
WL.3.RBPK													
ELA.3.26													
ELA.3.27													
WL.3.RW													
ELA.3.29													
WL.3.CSE													
ELA.3.36: [1]	17:(6)[1]	18:(6)[1]	19:(10)[1]	37:(6)[1]	39:(6)[1]	40:(6)[1]	41:(12)[1]						
ELA.3.37: [1]	19:(2)[1]	20:(6)[1]	38:(12)[1]	41:(12)[1]									
WL.3.KL													
ELA.3.38: [2]	42:(18)[3]												
WL.3.VAU													
ELA.3.39: [2]	2:(1)[2]												
ELA.3.40													
ELA.3.41													

Table 3.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
CONF WV Grade 3B1 ELA

Low	Medium	High
3.6	10.8	18

1 10199-4978	Exact:3	Partial:2	Minimal:1
2 10199-4986	Exact:6		
3 10199-4987	Exact:6		
4 10199-4989	Exact:5	Partial:1	
5 10199-4992	Exact:6		
6 10199-4993	Exact:6		
7 10199-11976	Exact:5	Partial:1	
8 10199-11978	Exact:6		
9 10199-11979	Exact:3	Partial:3	
10 10199-11986	Exact:6		
11 10199-11987	Exact:5	Partial:1	
12 10199-11988	Exact:2	Partial:4	
13 10199-11989	Exact:6		
14 10199-11991	Exact:6		
15 10199-13975	Exact:6		
16 10199-13976	Exact:6		
17 10199-1173	Exact:6		
18 10199-1181	Exact:6		
19 10199-1185	Exact:12		
20 10199-1191	Exact:6		
21 REP10199-6897	Exact:6		
22 REP10199-6899	Exact:6		
23 REP10199-6907	Exact:5	Partial:1	
24 REP10199-6911	Exact:6		
25 10199-6913	Exact:6		
26 REP10199-6914	Exact:6		
27 REP10199-6902	Exact:6		
28 REP10199-6915	Exact:6		
29 REP10199-6909	Exact:5	Partial:1	
30 REP10199-6586	Exact:4	Partial:2	

31 REP10199-6591	Exact:5	Partial:1	
32 REP10199-6592	Exact:6		
33 REP10199-6596	Exact:6		
34 REP10199-6598	Exact:4	Partial:2	
35 REP10199-6599	Exact:2	Minimal:3	Negligible:1
36 REP10199-6601	Exact:6		
37 REP10199-1245	Exact:6		
38 REP10199-1246	Exact:12		
39 REP10199-1248	Exact:6		
40 REP10199-1250	Exact:6		
41 CONVENTIONS10199-14161	Minimal:12		
42 ELABORATION10199-14161	Partial:18		
43 ORGANIZATION10199-14161	Partial:18		

Grade 3 Batch 2 ELA West Virginia

Table 3.1 Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 3 B2 Number of Assessment Items - 42

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.3 Grade 3 Literary Reading	3	8	2 3	5 3	62.5 37.5	17.83	0.41	YES
IR.3 Grade 3 Informational Reading	3	9	2 3	5 4	55.56 44.44	14	1.26	YES
WL.3 Grade 3 Writing and Language	7	15	1 2 3 4	2 5 7 1	13.33 33.33 46.67 6.67	28.17	5.64	YES
Total	13	32	1 2 3 4	2 15 14 1	6 47 44 3	60	6	

Table 3.2 Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 3 B2 Number of Assessment Items - 42

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.3 Grade 3 Literary Reading	3	8	17.83	0.41	27.83	17	70.32	17	1.85	3	YES
IR.3 Grade 3 Informational Reading	3	9	14	1.26	29.45	13	68.23	13	2.32	4	YES
WL.3 Grade 3 Writing and Language	7	15	28.17	5.64	0	0	80.29	10	19.71	10	YES
Total	13	32	60	6	15.28	4.4	74.17	8.4	10.56	4.5	

Table 3.3

Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers

WV ELA 2019 Grade 3 B2

Number of Assessment Items - 42

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.3 Grade 3 Literary Reading	3	8	17.83	0.41	7.67	0.52	95.83	6.45	YES	39	2	0.81	0.04	YES
IR.3 Grade 3 Informational Reading	3	9	14	1.26	7.67	1.03	85.19	11.48	YES	30	2	0.75	0.1	YES
WL.3 Grade 3 Writing and Language	7	15	28.17	5.64	5.17	0.98	34.44	6.55	NO	32	3	0.7	0.1	YES
Total	13	32	60	6	6.8	1.44	71.82	33		34	5	0.75	0.06	

Table 3.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers

WV ELA 2019 Grade 3 B2 Number of Assessment Items - 42

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.3 Grade 3 Literary Reading	YES	YES	YES	YES
IR.3 Grade 3 Informational Reading	YES	YES	YES	YES
WL.3 Grade 3 Writing and Language	YES	YES	NO	YES

Table 3.5 *Depth-of-Knowledge Levels by Item and Reviewers*
Intraclass Correlation WV ELA 2019 Grade 3 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6	Reviewer 7
1	2		2	2	2	2	2
2	2		2	2	2	2	2
3	1		2	2	2	2	1
4	2		2	2	2	2	2
5	2		2	3	3	3	2
6	1		1	1	1	1	1
7	2		3	3	3	2	2
8	2		2	2	2	2	2
9	3		3	3	3	3	2
10	2		2	2	2	3	2
11	2		2	2	2	2	2
12	2		2	2	3	2	2
13	2		1	2	2	2	1
14	1		1	1	1	1	1
15	1		1	1	1	1	1
16	1		1	1	1	1	1
17	1		1	1	1	1	1
18	2		2	2	2	2	2
19	2		3	3	3	3	2
20	2		2	3	2	2	2
21	3		2	2	3	2	2
22	2		2	2	2	2	2
23	2		2	2	2	1	2
24	3		3	3	2	3	2
25	2		3	3	3	3	2
26	2		1	2	2	1	2
27	2		2	3	3	3	2
28	2		2	2	2	2	2
29	2		1	2	2	2	2
30	2		3	2	2	2	2
31	2		2	1	2	1	2

32	2		1	1	2	1	2
33	2		2	2	2	2	2
34	3		3	3	2	3	2
35	2		2	1	2	1	2
36	1		1	1	1	1	1
37	1		2	1	1	1	1
38	1		1	1	1	1	1
39	1		1	1	1	1	1
40	1		2	1	1	1	1
41	3		3	3	3	3	3
42	3		3	3	3	3	3

Intraclass correlation - .9538

Pairwise Comparison - 0.74

Table 3.6

DOK Levels and Objectives Code by Each Reviewer

WV ELA 2019 Grade 3 B2

Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.3.2			2	ELA.3.2			2	ELA.3.2			2	ELA.3.2			2	ELA.3.2			2	ELA.3.2		
2	2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.1	ELA.3.7		2	ELA.3.7			2	ELA.3.7		
3	2	ELA.3.1			2	ELA.3.9			1	ELA.3.9			1	ELA.3.1			2	ELA.3.9			2	ELA.3.9		
4	2	ELA.3.8			2	ELA.3.8			2	ELA.3.8			2	ELA.3.8			2	ELA.3.8			2	ELA.3.8		
5	2	ELA.3.3			3	ELA.3.3			2	ELA.3.3			2	ELA.3.3			3	ELA.3.3			3	ELA.3.3		
6	1	ELA.3.3			1	ELA.3.1			1	ELA.3.1			1	ELA.3.1			1	ELA.3.1			1	ELA.3.1		
7	3	ELA.3.1 3			2	ELA.3.1 3			2	ELA.3.1 3			2	ELA.3.1 3			3	ELA.3.1 3			3	ELA.3.1 3		
8	2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.4 0			2	ELA.3.1 0			2	ELA.3.7		
9	3	ELA.3.1 5			3	ELA.3.1 5			3	ELA.3.1 5			2	ELA.3.1 1	ELA.3.1 5		3	ELA.3.1 5			3	ELA.3.1 1		
10	2	ELA.3.1 5			3	ELA.3.1 5			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4		
11	2	ELA.3.5			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.1 6		
12	2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			3	ELA.3.4			2	ELA.3.4		
13	1	ELA.3.1 1			2	ELA.3.1 1			2	ELA.3.1 1			1	ELA.3.1 1			2	ELA.3.1 1			2	ELA.3.1 1		
14	1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6		
15	1	ELA.3.3 6	ELA.3.3 7		1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.6			1	ELA.3.3 6		
16	1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6		
17	1	ELA.3.3 7			1	ELA.3.3 6			1	ELA.3.3 7			1	ELA.3.3 7			1	ELA.3.3 7			1	ELA.3.3 7		
18	2	ELA.3.2			2	ELA.3.8			2	ELA.3.1			2	ELA.3.3			2	ELA.3.8			2	ELA.3.3		
19	3	ELA.3.1 4			3	ELA.3.1 4			2	ELA.3.1 4			2	ELA.3.1 4			3	ELA.3.1 4			3	ELA.3.1 4		
20	2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			3	ELA.3.1 4		

21	2	ELA.3.1			2	ELA.3.1			3	ELA.3.3			2	ELA.3.3			3	ELA.3.3		
22	2	ELA.3.1 3			2	ELA.3.8			2	ELA.3.8			2	ELA.3.1			2	ELA.3.8		
23	2	ELA.3.1			1	ELA.3.1			2	ELA.3.1			2	ELA.3.1			2	ELA.3.1		
24	3	ELA.3.2			3	ELA.3.3			3	ELA.3.1 4			2	ELA.3.3			2	ELA.3.1		
25	3	ELA.3.1 3			3	ELA.3.1 4			2	ELA.3.1 4			2	ELA.3.1 4			3	ELA.3.3		
26	1	ELA.3.7			1	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7		
27	2	ELA.3.1 3			3	ELA.3.1 3			2	ELA.3.1 3			2	ELA.3.1 3			3	ELA.3.1 3		
28	2	ELA.3.4			2	ELA.3.1 6			2	ELA.3.1 6			2	ELA.3.4			2	ELA.3.4		
29	1	ELA.3.5			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4		
30	3	ELA.3.1 7			2	ELA.3.5			2	ELA.3.5			2	ELA.3.5			2	ELA.3.5		
31	2	ELA.3.6			1	ELA.3.4			2	ELA.3.6			2	ELA.3.4			2	ELA.3.4		
32	1	ELA.3.1 0			1	ELA.3.1 0			2	ELA.3.1 0			2	ELA.3.1 0			2	ELA.3.1 0		
33	2	ELA.3.1 2			2	ELA.3.1 2			2	ELA.3.1 2			2	ELA.3.4			2	ELA.3.1 2		
34	3	ELA.3.1 7			3	ELA.3.1 7			3	ELA.3.1 7			2	ELA.3.1 7			2	ELA.3.1 7		
35	2	ELA.3.1 0			1	ELA.3.1 0			2	ELA.3.1 0			2	ELA.3.1 0			2	ELA.3.1 0		
36	1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6		
37	2	ELA.3.6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6			1	ELA.3.3 6		
38	1	ELA.3.3 7			1	ELA.3.3 6			1	ELA.3.3 7			1	ELA.3.3 6			1	ELA.3.3 7		
39	1	ELA.3.3 7			1	ELA.3.3 6			1	ELA.3.3 7			1	ELA.3.3 7			1	ELA.3.3 6		
40	2	ELA.3.3 6	ELA.3.3 7		1	ELA.3.3 7	ELA.3.3 6		1	ELA.3.3 6	ELA.3.3 7		1	ELA.3.3 6	ELA.3.3 7		1	ELA.3.3 6	ELA.3.3 7	
41	3	ELA.3.3 8	ELA.3.2 0		3	ELA.3.3 8			3	ELA.3.2 0	ELA.3.3 8		3	ELA.3.2 0	ELA.3.3 8		3	ELA.3.2 0	ELA.3.3 8	
42	3	ELA.3.2 0	ELA.3.3 8	ELA.3.2 3	3	ELA.3.2 1	ELA.3.3 8		3	ELA.3.2 0	ELA.3.2 3		3	ELA.3.2 0	ELA.3.2 3		3	ELA.3.2 0	ELA.3.2 3	
Objective Pairwise Comparison: 0.67																				
Standard Pairwise Comparison: 0.97																				

Table 3.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV ELA 2019 Grade 3 B2

	Low			Medium			High		
	0			12.6				21	
LR.3									
LR.3.KID									
ELA.3.1	2(1)	3(2)	6(5)	18(1)	21(2)	22(1)	23(6)	24(1)	
ELA.3.2	24(1)	25(1)	18(1)	1(6)					
ELA.3.3	6(1)	5(6)	18(2)	21(4)	25(1)	24(2)			
LR.3.CS									
ELA.3.7	26(6)	20(5)	8(4)	2(6)					
ELA.3.8	4(6)	22(4)	18(2)						
ELA.3.9	3(4)								
LR.3.IKI									
ELA.3.13	7(6)	22(1)	27(6)	25(1)					
ELA.3.14	25(3)	24(2)	20(1)	19(6)					
IR.3									
IR.3.KID									
ELA.3.4	10(4)	11(4)	29(4)	28(4)	12(6)	31(4)	33(1)		
ELA.3.5	30(5)	29(1)	11(1)						
ELA.3.6	15(2)	31(2)	37(2)						
IR.3.CS									
ELA.3.10	35(6)	32(6)	8(1)						
ELA.3.11	13(6)	9(2)							
ELA.3.12	33(5)								
IR.3.IKI									
ELA.3.15	10(2)	9(5)							

ELA.3.16	11(1)	29(1)	28(2)						
ELA.3.17	34(6)	30(1)							
WL.3									
WL.3.TTP									
ELA.3.20	42(12)	41(12)							
ELA.3.21	42(9)	41(6)							
ELA.3.22									
WL.3.PDW									
ELA.3.23	42(15)								
ELA.3.24									
ELA.3.25									
WL.3.RBPK									
ELA.3.26									
ELA.3.27									
WL.3.RW									
ELA.3.29									
WL.3.CSE									
ELA.3.36	39(3)	36(6)	37(10)	38(2)	14(6)	15(12)	16(6)	17(1)	40(12)
ELA.3.37	17(5)	38(4)	39(3)	40(12)	15(2)				
WL.3.KL									
ELA.3.38	42(9)	41(21)							
WL.3.VAU									
ELA.3.39									
ELA.3.40	8(1)								
ELA.3.41									

Table 3.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 3 B2

Low		Medium		High	
4.2		12.6		21	
1	10199-4959	ELA.3.2:6			
2	10199-4966	ELA.3.1:1	ELA.3.7:6		
3	10199-4968	ELA.3.1:2	ELA.3.9:4		
4	10199-4969	ELA.3.8:6			
5	10199-4970	ELA.3.3:6			
6	10199-4972	ELA.3.1:5	ELA.3.3:1		
7	10199-7076	ELA.3.13:6			
8	REP10199-12211	ELA.3.7:4	ELA.3.10:1	ELA.3.40:1	
9	REP10199-12212	ELA.3.11:2	ELA.3.15:5		
10	REP10199-12213	ELA.3.4:4	ELA.3.15:2		
11	10199-12215	ELA.3.4:4	ELA.3.5:1	ELA.3.16:1	
12	REP10199-12216	ELA.3.4:6			
13	REP10199-14011	ELA.3.11:6			
14	10199-3051	ELA.3.36:6			
15	10199-3052	ELA.3.6:2	ELA.3.36:12	ELA.3.37:2	
16	10199-3053	ELA.3.36:6			
17	10199-3060	ELA.3.36:1	ELA.3.37:5		
18	10199-6833	ELA.3.1:1	ELA.3.2:1	ELA.3.3:2	ELA.3.8:2
19	10199-6835	ELA.3.14:6			
20	10199-6836	ELA.3.7:5	ELA.3.14:1		
21	10199-6837	ELA.3.1:2	ELA.3.3:4		
22	10199-6838	ELA.3.1:1	ELA.3.8:4	ELA.3.13:1	
23	10199-6841	ELA.3.1:6			
24	10199-6839	ELA.3.1:1	ELA.3.2:1	ELA.3.3:2	ELA.3.14:2
25	10199-6842	ELA.3.2:1	ELA.3.3:1	ELA.3.13:1	ELA.3.14:3
26	10199-6840	ELA.3.7:6			
27	10199-7077	ELA.3.13:6			
28	10199-6844	ELA.3.4:4	ELA.3.16:2		
29	10199-6845	ELA.3.4:4	ELA.3.5:1	ELA.3.16:1	
30	10199-6846	ELA.3.5:5	ELA.3.17:1		

31 10199-6847	ELA.3.4:4	ELA.3.6:2	
32 10199-6849	ELA.3.10:6		
33 10199-6850	ELA.3.4:1	ELA.3.12:5	
34 10199-6854	ELA.3.17:6		
35 10199-6856	ELA.3.10:6		
36 10199-1837	ELA.3.36:6		
37 10199-1838	ELA.3.6:2	ELA.3.36:10	
38 10199-1844	ELA.3.36:2	ELA.3.37:4	
39 10199-1848	ELA.3.36:3	ELA.3.37:3	
40 CONVENTIONS10199-14162	ELA.3.36:12	ELA.3.37:12	
41 ELABORATION10199-14162	ELA.3.20:12	ELA.3.21:6	ELA.3.38:21
42 ORGANIZATION10199-14162	ELA.3.20:12	ELA.3.21:9	ELA.3.23:15
			ELA.3.38:9

Table 3.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 3 B2

Low DOK		Matched DOK		High DOK

LR.3									
LR.3.KID									
ELA.3.1: [2]	2:(1)[2]	3:(2)[2]	6:(5)[1]	18:(1)[2]	21:(2)[2]	22:(1)[2]	23:(6)[2]	24:(1)[2]	
ELA.3.2: [2]	1:(6)[2]	18:(1)[2]	24:(1)[3]	25:(1)[3]					
ELA.3.3: [3]	5:(6)[2]	6:(1)[1]	18:(2)[2]	21:(4)[2]	24:(2)[2]	25:(1)[3]			
LR.3.CS									
ELA.3.7: [2]	2:(6)[2]	8:(4)[2]	20:(5)[2]	26:(6)[2]					
ELA.3.8: [2]	4:(6)[2]	18:(2)[2]	22:(4)[2]						
ELA.3.9: [2]	3:(4)[2]								
LR.3.IKI									
ELA.3.13: [3]	7:(6)[2]	22:(1)[2]	25:(1)[3]	27:(6)[2]					
ELA.3.14: [3]	19:(6)[3]	20:(1)[3]	24:(2)[3]	25:(3)[2]					
IR.3									
IR.3.KID									
ELA.3.4: [2]	10:(4)[2]	11:(4)[2]	12:(6)[2]	28:(4)[2]	29:(4)[2]	31:(4)[2]	33:(1)[2]		
ELA.3.5: [2]	11:(1)[2]	29:(1)[1]	30:(5)[2]						
ELA.3.6: [3]	15:(2)[1]	31:(2)[2]	37:(2)[2]						
IR.3.CS									
ELA.3.10: [2]	8:(1)[2]	32:(6)[2]	35:(6)[2]						
ELA.3.11: [2]	9:(2)[2]	13:(6)[2]							
ELA.3.12: [3]	33:(5)[2]								
IR.3.IKI									
ELA.3.15: [3]	9:(5)[3]	10:(2)[2]							
ELA.3.16: [2]	11:(1)[2]	28:(2)[2]	29:(1)[2]						
ELA.3.17: [3]	30:(1)[3]	34:(6)[3]							
WL.3									
WL.3.TTP									
ELA.3.20: [3]	41:(12)[3]	42:(12)[3]							

ELA.3.21: [3]	41:(6)[3]	42:(9)[3]							
ELA.3.22									
WL.3.PDW									
ELA.3.23: [3]	42:(15)[3]								
ELA.3.24									
ELA.3.25									
WL.3.RBPK									
ELA.3.26									
ELA.3.27									
WL.3.RW									
ELA.3.29									
WL.3.CSE									
ELA.3.36: [1]	14:(6)[1]	15:(12)[1]	16:(6)[1]	17:(1)[1]	36:(6)[1]	37:(10)[1]	38:(2)[1]	39:(3)[1]	40:(12)[1]
ELA.3.37: [1]	15:(2)[1]	17:(5)[1]	38:(4)[1]	39:(3)[1]	40:(12)[1]				
WL.3.KL									
ELA.3.38: [2]	41:(21)[3]	42:(9)[3]							
WL.3.VAU									
ELA.3.39									
ELA.3.40: [2]	8:(1)[2]								
ELA.3.41									

Table 3.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 3B2 ELA

Low		Medium		High	
3.6		10.8		18	
1	10199-4959	Exact:6			
2	10199-4966	Exact:6			
3	10199-4968	Exact:4	Partial:2		
4	10199-4969	Exact:6			
5	10199-4970	Exact:6			
6	10199-4972	Exact:1	Partial:5		
7	10199-7076	Exact:6			
8	REP10199-12211	Exact:6			
9	REP10199-12212	Exact:6			
10	REP10199-12213	Exact:3	Partial:2	Minimal:1	
11	10199-12215	Exact:6			
12	REP10199-12216	Exact:1	Partial:5		
13	REP10199-14011	Exact:6			
14	10199-3051	Exact:6			
15	10199-3052	Exact:12			
16	10199-3053	Exact:6			
17	10199-3060	Exact:6			
18	10199-6833	Exact:5	Minimal:1		
19	10199-6835	Exact:6			
20	10199-6836	Exact:6			
21	10199-6837	Exact:6			
22	10199-6838	Exact:5	Partial:1		
23	10199-6841	Exact:6			
24	10199-6839	Exact:5	Partial:1		
25	10199-6842	Exact:6			
26	10199-6840	Exact:6			
27	10199-7077	Exact:6			
28	10199-6844	Exact:6			
29	10199-6845	Exact:6			
30	10199-6846	Exact:6			

31 10199-6847	Exact:5	Partial:1	
32 10199-6849	Exact:6		
33 10199-6850	Exact:5	Partial:1	
34 10199-6854	Exact:6		
35 10199-6856	Exact:6		
36 10199-1837	Exact:6		
37 10199-1838	Exact:12		
38 10199-1844	Exact:6		
39 10199-1848	Exact:6		
40 CONVENTIONS10199-14162	Minimal:12		
41 ELABORATION10199-14162	Partial:18		
42 ORGANIZATION10199-14162	Partial:15	Partial:3	

Grade 3 Batch 4 ELA West Virginia

Table 3.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 3 B4 Number of Assessment Items - 43*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.3 Grade 3 Literary Reading	3	8	2 3	5 3	62.5 37.5	18	1.1	YES
IR.3 Grade 3 Informational Reading	3	9	2 3	5 4	55.56 44.44	14.5	0.55	YES
WL.3 Grade 3 Writing and Language	7	15	1 2 3 4	2 5 7 1	13.33 33.33 46.67 6.67	26	2.76	YES
Total	13	32	1 2 3 4	2 15 14 1	6 47 44 3	58.5	2.59	

Table 3.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 3 B4 Number of Assessment Items - 43*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.3 Grade 3 Literary Reading	3	8	18	1.1	26.57	13	69.08	16	4.35	7	YES
IR.3 Grade 3 Informational Reading	3	9	14.5	0.55	22.22	25	75.4	24	2.38	6	YES
WL.3 Grade 3 Writing and Language	7	15	26	2.76	0	0	85.58	10	14.42	10	YES
Total	13	32	58.5	2.59	13.68	10.1	77.78	12.1	8.55	3.7	

Table 3.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers
WV ELA 2019 Grade 3 B4 Number of Assessment Items - 43

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.3 Grade 3 Literary Reading	3	8	18	1.1	6.83	0.41	85.42	5.1	YES	38	1	0.71	0.04	YES
IR.3 Grade 3 Informational Reading	3	9	14.5	0.55	6.17	0.41	68.52	4.54	YES	31	1	0.83	0.04	YES
WL.3 Grade 3 Writing and Language	7	15	26	2.76	4.83	0.41	32.22	2.72	NO	31	1	0.74	0.08	YES
Total	13	32	58.5	2.59	5.9	1.02	62.05	27		33	4	0.76	0.06	

Table 3.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers
WV ELA 2019 Grade 3 B4 Number of Assessment Items - 43

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.3 Grade 3 Literary Reading	YES	YES	YES	YES
IR.3 Grade 3 Informational Reading	YES	YES	YES	YES
WL.3 Grade 3 Writing and Language	YES	YES	NO	YES

Table 3.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 3 B4

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6	Reviewer 7
1	2		2	2	2	2	2
2	3		3	3	3	3	2
3	2		2	2	2	2	2
4	2		2	2	2	2	2
5	2		1	2	2	2	1
6	2		2	2	2	1	2
7	2		2	2	2	2	2
8	2		2	2	2	2	2
9	2		2	1	2	2	2
10	2		2	3	2	3	2
11	2		3	3	3	3	2
12	2		3	3	3	3	2
13	2		2	2	3	3	2
14	2		2	3	2	1	2
15	2		2	2	2	1	2
16	2		2	2	2	2	2
17	1		1	1	1	1	1
18	1		2	1	1	1	1
19	1		2	1	1	1	1
20	1		1	1	1	1	1
21	2		2	1	2	1	2
22	2		2	2	2	1	2
23	2		2	2	2	2	2
24	2		1	1	2	1	2
25	2		2	3	2	2	2
26	3		3	3	3	2	2
27	3		3	3	3	2	2
28	2		1	3	2	1	2
29	3		3	3	3	2	2
30	2		2	2	2	1	2

31	3		2	3	2	2	2
32	3		2	2	3	2	2
33	2		2	2	2	1	2
34	2		2	2	2	1	2
35	2		2	2	3	2	2
36	2		2	2	2	1	2
37	1		1	1	1	1	1
38	1		1	1	1	1	1
39	1		1	1	1	1	1
40	1		1	1	1	1	1
41	1		1	1	1	1	1
42	3		3	3	3	3	3
43	3		3	3	3	3	3

Intraclass correlation - .9456

Pairwise Comparison - 0.72

Table 3.6
DOK Levels and Objectives Code by Each Reviewer
WV ELA 2019 Grade 3 B4

Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.3.7			2	ELA.3.7			2	ELA.3.10			2	ELA.3.40			2	ELA.3.10			2	ELA.3.10		
2	3	ELA.3.15			3	ELA.3.15			3	ELA.3.15			2	ELA.3.15	ELA.3.11		3	ELA.3.15			3	ELA.3.11		
3	2	ELA.3.15			2	ELA.3.15			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4		
4	2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4		
5	1	ELA.3.11			2	ELA.3.11			2	ELA.3.11			1	ELA.3.11			2	ELA.3.11			2	ELA.3.11		
6	2	ELA.3.5			1	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4		
7	2	ELA.3.13			2	ELA.3.8			2	ELA.3.1			2	ELA.3.1			2	ELA.3.8			2	ELA.3.8		
8	2	ELA.3.2			2	ELA.3.2			2	ELA.3.2			2	ELA.3.2			2	ELA.3.2			2	ELA.3.2		
9	2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			1	ELA.3.7		
10	2	ELA.3.1			3	ELA.3.3			2	ELA.3.8			2	ELA.3.8			2	ELA.3.8			3	ELA.3.3		
11	3	ELA.3.14			3	ELA.3.14			2	ELA.3.14			2	ELA.3.14			3	ELA.3.14			3	ELA.3.14		
12	3	ELA.3.3			3	ELA.3.3			2	ELA.3.3			2	ELA.3.3			3	ELA.3.3			3	ELA.3.3		
13	2	ELA.3.1			3	ELA.3.3			2	ELA.3.1			2	ELA.3.3			3	ELA.3.3			2	ELA.3.3		
14	2	ELA.3.1			1	ELA.3.1			2	ELA.3.1			2	ELA.3.1			2	ELA.3.1			3	ELA.3.1		
15	2	ELA.3.9			1	ELA.3.9			2	ELA.3.1			2	ELA.3.1			2	ELA.3.1			2	ELA.3.9		
16	2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7			2	ELA.3.7		
17	1	ELA.3.37			1	ELA.3.37			1	ELA.3.37			1	ELA.3.37			1	ELA.3.37			1	ELA.3.36		
18	2	ELA.3.36			1	ELA.3.36			1	ELA.3.36			1	ELA.3.36			1	ELA.3.36			1	ELA.3.36		
19	2	ELA.3.36			1	ELA.3.36			1	ELA.3.36			1	ELA.3.36			1	ELA.3.36			1	ELA.3.36		
20	1	ELA.3.37			1	ELA.3.37			1	ELA.3.37			1	ELA.3.37			1	ELA.3.37			1	ELA.3.37		
21	2	ELA.3.4			1	ELA.3.4			2	ELA.3.4			2	ELA.3.4			2	ELA.3.4			1	ELA.3.1		

22	2	ELA.3.4		1	ELA.3.4		2	ELA.3.4		2	ELA.3.4		2	ELA.3.4		2	ELA.3.4		
23	2	ELA.3.5		2	ELA.3.5		2	ELA.3.5		2	ELA.3.5		2	ELA.3.5		2	ELA.3.5		
24	1	ELA.3.1 0		1	ELA.3.1 0		2	ELA.3.1 0		2	ELA.3.1 0		2	ELA.3.1 0		1	ELA.3.1 0		
25	2	ELA.3.1 1		2	ELA.3.1 1		2	ELA.3.1 1		2	ELA.3.5		2	ELA.3.1 1		3	ELA.3.6		
26	3	ELA.3.1 5		2	ELA.3.1 5		3	ELA.3.1 5		2	ELA.3.1 5		3	ELA.3.1 5		3	ELA.3.1 5		
27	3	ELA.3.1 7		2	ELA.3.1 7		3	ELA.3.1 7		2	ELA.3.1 7		3	ELA.3.1 7		3	ELA.3.1 7		
28	1	ELA.3.1 0		1	ELA.3.1 0		2	ELA.3.1 0		2	ELA.3.1 0		2	ELA.3.1 0		3	ELA.3.1 0		
29	3	ELA.3.1 7		2	ELA.3.1 7		3	ELA.3.1 7		2	ELA.3.1 7		3	ELA.3.1 7		3	ELA.3.1 7		
30	2	ELA.3.3		1	ELA.3.3		2	ELA.3.3		2	ELA.3.3		2	ELA.3.3		2	ELA.3.3		
31	2	ELA.3.1		2	ELA.3.3		3	ELA.3.3		2	ELA.3.1		2	ELA.3.2		3	ELA.3.2		
32	2	ELA.3.3		2	ELA.3.9		3	ELA.3.3		2	ELA.3.3		3	ELA.3.3		2	ELA.3.3		
33	2	ELA.3.7		1	ELA.3.7		2	ELA.3.7		2	ELA.3.7		2	ELA.3.7		2	ELA.3.7		
34	2	ELA.3.9		1	ELA.3.9		2	ELA.3.9		2	ELA.3.1		2	ELA.3.9		2	ELA.3.9		
35	2	ELA.3.3		2	ELA.3.3		2	ELA.3.3		2	ELA.3.3		3	ELA.3.3		2	ELA.3.3		
36	2	ELA.3.7		1	ELA.3.7		2	ELA.3.7		2	ELA.3.7		2	ELA.3.7		2	ELA.3.7		
37	1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 6		1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		
38	1	ELA.3.3 6	ELA.3.3 7	1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 6		1	ELA.3.3 7		
39	1	ELA.3.3 7		1	ELA.3.3 6		1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		
40	1	ELA.3.3 6		1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		1	ELA.3.3 7		
41	1	ELA.3.3 6	ELA.3.3 7	1	ELA.3.3 7	ELA.3.3 6	1	ELA.3.3 6	ELA.3.3 7	1	ELA.3.3 6	ELA.3.3 7	1	ELA.3.3 6	ELA.3.3 7	1	ELA.3.3 6	ELA.3.3 7	
42	3	ELA.3.2 1	ELA.3.3 8	3	ELA.3.3 8		3	ELA.3.2 1	ELA.3.2 3	3	ELA.3.2 1	ELA.3.8	3	ELA.3.2 1	ELA.3.3 8	3	ELA.3.2 1	ELA.3.3 8	
43	3	ELA.3.2 1	ELA.3.2 3	3	ELA.3.3 8	ELA.3.2 1	3	ELA.3.2 1	ELA.3.3 8	3	ELA.3.2 1	ELA.3.2 3	3	ELA.3.2 3	ELA.3.2 1	3	ELA.3.2 1	ELA.3.2 3	
Objective Pairwise Comparison: 0.75																			
Standard Pairwise Comparison: 0.97																			

Table 3.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV ELA 2019 Grade 3 B4

	Low		Medium			High		
	0		10.8				18	
LR.3								
LR.3.KID								
ELA.3.1	10(1)	7(2)	14(6)	15(3)	21(1)	13(2)	31(2)	34(1)
ELA.3.2	31(2)	8(6)						
ELA.3.3	12(6)	10(2)	13(4)	30(6)	31(2)	32(5)	35(6)	
LR.3.CS								
ELA.3.7	36(6)	33(6)	16(6)	9(6)	1(2)			
ELA.3.8	10(3)	7(3)	42(3)					
ELA.3.9	15(3)	34(5)	32(1)					
LR.3.IKI								
ELA.3.13	7(1)							
ELA.3.14	11(6)							
IR.3								
IR.3.KID								
ELA.3.4	4(6)	3(4)	6(5)	21(5)	22(6)			
ELA.3.5	23(6)	25(1)	6(1)					
ELA.3.6	25(1)							
IR.3.CS								
ELA.3.10	28(6)	24(6)	1(3)					
ELA.3.11	5(6)	25(4)	2(2)					
ELA.3.12								
IR.3.IKI								
ELA.3.15	26(6)	2(5)	3(2)					

ELA.3.16								
ELA.3.17	27(6)	29(6)						
WL.3								
WL.3.TTP								
ELA.3.20								
ELA.3.21	42(15)	43(18)						
ELA.3.22								
WL.3.PDW								
ELA.3.23	42(3)	43(12)						
ELA.3.24								
ELA.3.25								
WL.3.RBPK								
ELA.3.26								
ELA.3.27								
WL.3.RW								
ELA.3.29								
WL.3.CSE								
ELA.3.36	38(10)	39(1)	40(1)	37(1)	18(6)	19(6)	17(1)	41(12)
ELA.3.37	17(5)	20(12)	37(5)	40(5)	39(5)	38(4)	41(12)	
WL.3.KL								
ELA.3.38	42(12)	43(9)						
WL.3.VAU								
ELA.3.39								
ELA.3.40	1(1)							
ELA.3.41								

Table 3.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 3 B4

Low		Medium		High	
3.6		10.8		18	
1	REP10199-12211	ELA.3.7:2	ELA.3.10:3	ELA.3.40:1	
2	REP10199-12212	ELA.3.11:2	ELA.3.15:5		
3	REP10199-12213	ELA.3.4:4	ELA.3.15:2		
4	REP10199-12216	ELA.3.4:6			
5	REP10199-14011	ELA.3.11:6			
6	10199-14015	ELA.3.4:5	ELA.3.5:1		
7	10199-12040	ELA.3.1:2	ELA.3.8:3	ELA.3.13:1	
8	10199-12042	ELA.3.2:6			
9	10199-12044	ELA.3.7:6			
10	10199-12046	ELA.3.1:1	ELA.3.3:2	ELA.3.8:3	
11	10199-12050	ELA.3.14:6			
12	10199-12051	ELA.3.3:6			
13	10199-12052	ELA.3.1:2	ELA.3.3:4		
14	10199-13919	ELA.3.1:6			
15	10199-13924	ELA.3.1:3	ELA.3.9:3		
16	10199-13927	ELA.3.7:6			
17	10199-2978	ELA.3.36:1	ELA.3.37:5		
18	10199-3028	ELA.3.36:6			
19	10199-3030	ELA.3.36:6			
20	10199-3032	ELA.3.37:12			
21	REP10199-6895	ELA.3.1:1	ELA.3.4:5		
22	REP10199-6897	ELA.3.4:6			
23	REP10199-6899	ELA.3.5:6			
24	REP10199-6907	ELA.3.10:6			
25	REP10199-6911	ELA.3.5:1	ELA.3.6:1	ELA.3.11:4	
26	REP10199-6912	ELA.3.15:6			
27	REP10199-6914	ELA.3.17:6			
28	REP10199-6915	ELA.3.10:6			
29	REP10199-6909	ELA.3.17:6			
30	REP10199-6586	ELA.3.3:6			

31 REP10199-6591	ELA.3.1:2	ELA.3.2:2	ELA.3.3:2	
32 REP10199-6592	ELA.3.3:5	ELA.3.9:1		
33 REP10199-6596	ELA.3.7:6			
34 REP10199-6598	ELA.3.1:1	ELA.3.9:5		
35 REP10199-6599	ELA.3.3:6			
36 REP10199-6601	ELA.3.7:6			
37 10199-2238	ELA.3.36:1	ELA.3.37:5		
38 10199-2240	ELA.3.36:10	ELA.3.37:4		
39 10199-2241	ELA.3.36:1	ELA.3.37:5		
40 10199-2242	ELA.3.36:1	ELA.3.37:5		
41 CONVENTIONS10199-14161	ELA.3.36:12	ELA.3.37:12		
42 ELABORATION10199-14161	ELA.3.8:3	ELA.3.21:15	ELA.3.23:3	ELA.3.38:12
43 ORGANIZATION10199-14161	ELA.3.21:18	ELA.3.23:12	ELA.3.38:9	

Table 3.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 3 B4

Low DOK		Matched DOK		High DOK

LR.3								
LR.3.KID								
ELA.3.1: [2]	7:(2)[2]	10:(1)[2]	13:(2)[2]	14:(6)[2]	15:(3)[2]	21:(1)[1]	31:(2)[2]	34:(1)[2]
ELA.3.2: [2]	8:(6)[2]	31:(2)[2]						
ELA.3.3: [3]	10:(2)[3]	12:(6)[3]	13:(4)[2]	30:(6)[2]	31:(2)[2]	32:(5)[2]	35:(6)[2]	
LR.3.CS								
ELA.3.7: [2]	1:(2)[2]	9:(6)[2]	16:(6)[2]	33:(6)[2]	36:(6)[2]			
ELA.3.8: [2]	7:(3)[2]	10:(3)[2]	42:(3)[3]					
ELA.3.9: [2]	15:(3)[2]	32:(1)[2]	34:(5)[2]					
LR.3.IKI								
ELA.3.13: [3]	7:(1)[2]							
ELA.3.14: [3]	11:(6)[3]							
IR.3								
IR.3.KID								
ELA.3.4: [2]	3:(4)[2]	4:(6)[2]	6:(5)[2]	21:(5)[2]	22:(6)[2]			
ELA.3.5: [2]	6:(1)[2]	23:(6)[2]	25:(1)[2]					
ELA.3.6: [3]	25:(1)[3]							
IR.3.CS								
ELA.3.10: [2]	1:(3)[2]	24:(6)[2]	28:(6)[2]					
ELA.3.11: [2]	2:(2)[2]	5:(6)[2]	25:(4)[2]					
ELA.3.12								
IR.3.IKI								
ELA.3.15: [3]	2:(5)[3]	3:(2)[2]	26:(6)[3]					
ELA.3.16								
ELA.3.17: [3]	27:(6)[3]	29:(6)[3]						
WL.3								
WL.3.TTP								
ELA.3.20								

ELA.3.21: [3]	42:(15)[3]	43:(18)[3]						
ELA.3.22								
WL.3.PDW								
ELA.3.23: [3]	42:(3)[3]	43:(12)[3]						
ELA.3.24								
ELA.3.25								
WL.3.RBPK								
ELA.3.26								
ELA.3.27								
WL.3.RW								
ELA.3.29								
WL.3.CSE								
ELA.3.36: [1]	17:(1)[1]	18:(6)[1]	19:(6)[1]	37:(1)[1]	38:(10)[1]	39:(1)[1]	40:(1)[1]	41:(12)[1]
ELA.3.37: [1]	17:(5)[1]	20:(12)[1]	37:(5)[1]	38:(4)[1]	39:(5)[1]	40:(5)[1]	41:(12)[1]	
WL.3.KL								
ELA.3.38: [2]	42:(12)[3]	43:(9)[3]						
WL.3.VAU								
ELA.3.39								
ELA.3.40: [2]	1:(1)[2]							
ELA.3.41								

Grade 4 Batch 1 ELA West Virginia

Table 4.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV 2019 G4B1ELA v2 Number of Assessment Items - 43*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.4 Grade 4 Literary Reading	3	8	2 3	6 2	75 25	17.5	0.55	YES
IR.4 Grade 4 Informational Reading	3	9	2 3	5 4	55.56 44.44	13.5	1.05	YES
WL.4 Grade 4 Writing and Language	7	16	1 2 3 4	1 6 8 1	6.25 37.5 50 6.25	28.83	2.14	YES
Total	13	33	1 2 3 4	1 17 14 1	3 52 42 3	59.83	2.32	

Table 4.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Six Reviewers WV 2019 G4B1ELA v2 Number of Assessment Items - 43*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.4 Grade 4 Literary Reading	3	8	17.5	0.55	16.34	9	76.85	13	6.81	7	YES
IR.4 Grade 4 Informational Reading	3	9	13.5	1.05	27.48	15	69.02	18	3.5	6	YES
WL.4 Grade 4 Writing and Language	7	16	28.83	2.14	23.06	11	61	6	15.94	6	YES
Total	13	33	59.83	2.32	21.73	8.4	67.69	9.5	10.58	4.8	

Table 4.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers
WV 2019 G4B1ELA v2 Number of Assessment Items - 43

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.4 Grade 4 Literary Reading	3	8	17.5	0.55	7.67	0.52	95.83	6.45	YES	37	2	0.83	0.01	YES
IR.4 Grade 4 Informational Reading	3	9	13.5	1.05	6.17	0.41	68.52	4.54	YES	29	2	0.79	0.03	YES
WL.4 Grade 4 Writing and Language	7	16	28.83	2.14	6.5	0.55	40.62	3.42	WEAK	33	2	0.72	0.02	YES
Total	13	33	59.83	2.32	6.8	0.79	68.32	28		33	4	0.78	0.06	

Table 4.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers
WV 2019 G4B1ELA v2 Number of Assessment Items - 43

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.4 Grade 4 Literary Reading	YES	YES	YES	YES
IR.4 Grade 4 Informational Reading	YES	YES	YES	YES
WL.4 Grade 4 Writing and Language	YES	YES	WEAK	YES

Table 4.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV 2019 G4B1ELA v2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6	Reviewer 7
1	1	1	1	1		1	1
2	2	1	1	1		1	1
3	1	1	1	1		1	1
4	1	1	1	1		1	1
5	2	2	2	2		2	2
6	3	2	1	2		2	2
7	3	2	3	2		2	2
8	2	1	2	2		2	2
9	2	1	2	2		2	2
10	2	2	2	2		2	2
11	2	3	2	2		2	2
12	2	2	2	2		2	2
13	2	2	2	2		2	2
14	2	1	2	2		2	1
15	2	3	3	3		3	3
16	2	2	2	2		2	3
17	2	1	2	2		2	2
18	2	3	3	2		3	3
19	2	2	2	2		2	2
20	3	3	2	2		3	2
21	2	2	2	2		2	2
22	1	1	2	1		2	2
23	2	2	2	2		2	3
24	2	3	2	2		2	2
25	1	1	1	2		2	2
26	1	1	2	2		2	1
27	2	1	2	2		2	2
28	2	2	2	2		3	2
29	3	3	3	3		3	3
30	2	2	2	2		3	2

31	2	2	2	2		2	2
32	2	2	2	2		2	3
33	2	2	2	2		2	2
34	1	1	2	2		2	2
35	3	3	2	2		3	3
36	2	3	2	2		3	3
37	2	1	3	2		3	2
38	1	1	1	1		1	1
39	1	1	1	1		1	1
40	1	1	1	1		1	1
41	2	1	1	2		1	1
42	3	3	3	3		3	3
43	3	3	3	3		3	3

Intraclass correlation - .9392

Pairwise Comparison - 0.72

Table 4.6
DOK Levels and Objectives Code by Each Reviewer
WV 2019 G4B1ELA v2

Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	ELA.4.3 7			1	ELA.4.4			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7		
2	2	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6		
3	1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6		
4	1	ELA.4.3 7			1	ELA.4.3 6			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7		
5	2	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5		
6	3	ELA.4.1 1			2	ELA.4.1 1			1	ELA.4.1 1			2	ELA.4.1 1			2	ELA.4.1 1			2	ELA.4.1 1		
7	3	ELA.4.1 5			2	ELA.4.1 5			3	ELA.4.1 5			2	ELA.4.1 5			2	ELA.4.1 5			2	ELA.4.1 5		
8	2	ELA.4.1 0			1	ELA.4.3 9			2	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.1 0		
9	2	ELA.4.1 0			1	ELA.4.3 9			2	ELA.4.1 0			2	ELA.4.3 9			2	ELA.4.1 0			2	ELA.4.1 0		
10	2	ELA.4.4			2	ELA.4.4			2	ELA.4.5			2	ELA.4.4			2	ELA.4.4			2	ELA.4.4		
11	2	ELA.4.1			3	ELA.4.1			2	ELA.4.1			2	ELA.4.1			2	ELA.4.1			2	ELA.4.1		
12	2	ELA.4.3			2	ELA.4.3			2	ELA.4.3			2	ELA.4.3			2	ELA.4.3			2	ELA.4.3		
13	2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2		
14	2	ELA.4.7			1	ELA.4.7			2	ELA.4.7			2	ELA.4.7			2	ELA.4.7			1	ELA.4.7		
15	2	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4		
16	2	ELA.4.9			2	ELA.4.9			2	ELA.4.9			2	ELA.4.9			2	ELA.4.9			3	ELA.4.9		
17	2	ELA.4.7			1	ELA.4.7			2	ELA.4.7			2	ELA.4.7			2	ELA.4.7			2	ELA.4.3 9		
18	2	ELA.4.1			3	ELA.4.1 4			3	ELA.4.1 4			2	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4		
19	2	ELA.4.3			2	ELA.4.1			2	ELA.4.1			2	ELA.4.1			2	ELA.4.3			2	ELA.4.3		
20	3	ELA.4.1 4			3	ELA.4.1 4			2	ELA.4.1 4			2	ELA.4.1 4			3	ELA.4.1 4			2	ELA.4.1 4		

21	2	ELA.4.1 1			2	ELA.4.8			2	ELA.4.8			2	ELA.4.8			2	ELA.4.1 1		
22	1	ELA.4.4			1	ELA.4.4			2	ELA.4.4			1	ELA.4.4			2	ELA.4.4		
23	2	ELA.4.4			2	ELA.4.4			2	ELA.4.4			2	ELA.4.4			2	ELA.4.4		
24	2	ELA.4.5			3	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5		
25	1	ELA.4.4			1	ELA.4.4			1	ELA.4.4			2	ELA.4.4			2	ELA.4.4		
26	1	ELA.4.1 0			1	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.1 0		
27	2	ELA.4.1 0			1	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.3 9			2	ELA.4.1 0		
28	2	ELA.4.1 7			2	ELA.4.1 7			2	ELA.4.1 7			2	ELA.4.1 7			3	ELA.4.1 7		
29	3	ELA.4.6			3	ELA.4.1 7			3	ELA.4.1 7			3	ELA.4.1 7			3	ELA.4.1 7		
30	2	ELA.4.4 0			2	ELA.4.4 0			2	ELA.4.4 0			2	ELA.4.4 0			3	ELA.4.4 0		
31	2	ELA.4.7			2	ELA.4.4 0			2	ELA.4.7			2	ELA.4.7			2	ELA.4.7		
32	2	ELA.4.1			2	ELA.4.1			2	ELA.4.1			2	ELA.4.1			2	ELA.4.1		
33	2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2		
34	1	ELA.4.7			1	ELA.4.7			2	ELA.4.7			2	ELA.4.7			2	ELA.4.7		
35	3	ELA.4.3			3	ELA.4.3			2	ELA.4.3			2	ELA.4.3			3	ELA.4.3		
36	2	ELA.4.1 3			3	ELA.4.1 3			2	ELA.4.1 3			2	ELA.4.1 3			3	ELA.4.1 3		
37	2	ELA.4.1 3			1	ELA.4.1 3			3	ELA.4.1 3			2	ELA.4.1 3			3	ELA.4.1 3		
38	1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7		
39	1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7		
40	1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6	ELA.4.3 7	
41	2	ELA.4.3 6	ELA.4.3 7		1	ELA.4.3 6	ELA.4.3 7		1	ELA.4.3 6	ELA.4.3 7		2	ELA.4.3 6	ELA.4.3 7		1	ELA.4.3 6	ELA.4.3 7	
42	3	ELA.4.2 1	ELA.4.3 8		3	ELA.4.3 8			3	ELA.4.2 1	ELA.4.3 8		3	ELA.4.2 1	ELA.4.3 8		3	ELA.4.2 1	ELA.4.3 8	
43	3	ELA.4.2 1	ELA.4.2 3	ELA.4.3 8	3	ELA.4.2 1	ELA.4.2 8		3	ELA.4.2 1	ELA.4.2 3		3	ELA.4.2 1	ELA.4.2 3		3	ELA.4.2 1	ELA.4.2 3	ELA.4.3 8
Objective Pairwise Comparison: 0.79																				
Standard Pairwise Comparison: 0.92																				

Table 4.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers) WV 2019 G4B1ELA v2*

	Low		Medium		High	
	0		10.8			18
LR.4						
LR.4.KID						
ELA.4.1	11(6)	19(3)	18(1)	32(6)	35(1)	
ELA.4.2	33(6)	13(6)				
ELA.4.3	19(3)	12(6)	35(5)			
LR.4.CS						
ELA.4.7	34(6)	31(5)	17(5)	14(6)		
ELA.4.8	21(4)					
ELA.4.9	16(6)					
LR.4.IKI						
ELA.4.13	36(6)	37(6)				
ELA.4.14	18(5)	20(6)	23(1)	15(6)		
IR.4						
IR.4.KID						
ELA.4.4	10(5)	1(1)	23(5)	22(6)	25(6)	
ELA.4.5	24(5)	5(6)	10(1)			
ELA.4.6	29(1)					
IR.4.CS						
ELA.4.10	9(4)	8(5)	26(6)	27(4)		
ELA.4.11	21(2)	6(6)				
ELA.4.12						
IR.4.IKI						
ELA.4.15	7(6)					

ELA.4.16						
ELA.4.17	24(1)	28(6)	29(5)			
WL.4						
WL.4.TTP						
ELA.4.20						
ELA.4.21	42(15)	43(18)				
ELA.4.22						
WL.4.PDW						
ELA.4.23	43(15)					
ELA.4.24						
ELA.4.25						
WL.4.RBPK						
ELA.4.26						
ELA.4.27						
ELA.4.28	43(3)					
WL.4.RW						
ELA.4.29						
WL.4.CSE						
ELA.4.36	41(12)	40(6)	3(6)	4(1)	2(12)	
ELA.4.37	4(5)	1(5)	38(6)	39(12)	41(12)	40(8)
WL.4.KL						
ELA.4.38	42(18)	43(6)				
WL.4.VAU						
ELA.4.39	8(1)	9(2)	27(2)	17(1)		
ELA.4.40	31(1)	30(6)				
ELA.4.41						

Table 4.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 WV 2019 G4B1ELA v2

Low	Medium	High
3.6	10.8	18

1 10199-1333		ELA.4.4:1	ELA.4.37:5
2 10199-1335		ELA.4.36:12	
3 10199-1338		ELA.4.36:6	
4 10199-1341	ELA.4.36:1	ELA.4.37:5	
5 REP10199-13913	ELA.4.5:6		
6 REP10199-13921	ELA.4.11:6		
7 REP10199-13923	ELA.4.15:6		
8 REP10199-13928	ELA.4.10:5	ELA.4.39:1	
9 REP10199-13929	ELA.4.10:4	ELA.4.39:2	
10 REP10199-13934	ELA.4.4:5	ELA.4.5:1	
11 10199-7908	ELA.4.1:6		
12 REP10199-7911	ELA.4.3:6		
13 REP10199-7912	ELA.4.2:6		
14 REP10199-7913	ELA.4.7:6		
15 REP10199-7914	ELA.4.14:6		
16 REP10199-7934	ELA.4.9:6		
17 REP10199-7935	ELA.4.7:5	ELA.4.39:1	
18 10199-7963	ELA.4.1:1	ELA.4.14:5	
19 REP10199-7964	ELA.4.1:3	ELA.4.3:3	
20 REP10199-7965	ELA.4.14:6		
21 10199-8995	ELA.4.8:4	ELA.4.11:2	
22 10199-12073	ELA.4.4:6		
23 10199-12075	ELA.4.14:1	ELA.4.4:5	
24 10199-12078	ELA.4.5:5	ELA.4.17:1	
25 10199-12084	ELA.4.4:6		
26 10199-12085	ELA.4.10:6		
27 10199-12086	ELA.4.10:4	ELA.4.39:2	
28 10199-12096	ELA.4.17:6		
29 10199-12097	ELA.4.6:1	ELA.4.17:5	
30 10199-12098	ELA.4.40:6		

31 REP10199-3350	ELA.4.7:5	ELA.4.40:1	
32 REP10199-3355	ELA.4.1:6		
33 REP10199-3479	ELA.4.2:6		
34 REP10199-3667	ELA.4.7:6		
35 REP10199-3686	ELA.4.1:1	ELA.4.3:5	
36 REP10199-7065	ELA.4.13:6		
37 REP10199-7066	ELA.4.13:6		
38 REP10199-2025	ELA.4.37:6		
39 REP10199-2029	ELA.4.37:12		
40 REP10199-2040	ELA.4.36:6	ELA.4.37:8	
41 CONVENTIONS10199-14163	ELA.4.36:12	ELA.4.37:12	
42 ELABORATION10199-14163	ELA.4.21:15	ELA.4.38:18	
43 ORGANIZATION10199-14163	ELA.4.21:18	ELA.4.23:15	ELA.4.28:3 ELA.4.38:6

Table 4.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV 2019 G4B1ELA v2

Low DOK		Matched DOK		High DOK

LR.4						
LR.4.KID						
ELA.4.1: [2]	11:(6)[2]	18:(1)[2]	19:(3)[2]	32:(6)[2]	35:(1)[3]	
ELA.4.2: [2]	13:(6)[2]	33:(6)[2]				
ELA.4.3: [2]	12:(6)[2]	19:(3)[2]	35:(5)[3]			
LR.4.CS						
ELA.4.7: [2]	14:(6)[2]	17:(5)[2]	31:(5)[2]	34:(6)[2]		
ELA.4.8: [2]	21:(4)[2]					
ELA.4.9: [2]	16:(6)[2]					
LR.4.IKI						
ELA.4.13: [3]	36:(6)[2]	37:(6)[2]				
ELA.4.14: [3]	15:(6)[3]	18:(5)[3]	20:(6)[2]	23:(1)[3]		
IR.4						
IR.4.KID						
ELA.4.4: [2]	1:(1)[1]	10:(5)[2]	22:(6)[2]	23:(5)[2]	25:(6)[2]	
ELA.4.5: [2]	5:(6)[2]	10:(1)[2]	24:(5)[2]			
ELA.4.6: [2]	29:(1)[3]					
IR.4.CS						
ELA.4.10: [2]	8:(5)[2]	9:(4)[2]	26:(6)[2]	27:(4)[2]		
ELA.4.11: [2]	6:(6)[2]	21:(2)[2]				
ELA.4.12						
IR.4.IKI						
ELA.4.15: [3]	7:(6)[2]					
ELA.4.16						
ELA.4.17: [3]	24:(1)[2]	28:(6)[2]	29:(5)[3]			
WL.4						

WL.4.TTP						
ELA.4.20						
ELA.4.21: [3]	42:(15)[3]	43:(18)[3]				
ELA.4.22						
WL.4.PDW						
ELA.4.23: [3]	43:(15)[3]					
ELA.4.24						
ELA.4.25						
WL.4.RBPK						
ELA.4.26						
ELA.4.27						
ELA.4.28: [3]	43:(3)[3]					
WL.4.RW						
ELA.4.29						
WL.4.CSE						
ELA.4.36: [2]	2:(12)[1]	3:(6)[1]	4:(1)[1]	40:(6)[1]	41:(12)[1]	
ELA.4.37: [1]	1:(5)[1]	4:(5)[1]	38:(6)[1]	39:(12)[1]	40:(8)[1]	41:(12)[1]
WL.4.KL						
ELA.4.38: [2]	42:(18)[3]	43:(6)[3]				
WL.4.VAU						
ELA.4.39: [2]	8:(1)[1]	9:(2)[2]	17:(1)[2]	27:(2)[2]		
ELA.4.40: [3]	30:(6)[2]	31:(1)[2]				
ELA.4.41						

Grade 4 Batch 2 ELA West Virginia

Table 4.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV 2019 G4B2ELA v2 Number of Assessment Items - 42*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.4 Grade 4 Literary Reading	3	8	2 3	6 2	75 25	16.17	0.75	YES
IR.4 Grade 4 Informational Reading	3	9	2 3	5 4	55.56 44.44	14.33	1.21	YES
WL.4 Grade 4 Writing and Language	7	16	1 2 3 4	1 6 8 1	6.25 37.5 50 6.25	28.5	1.87	YES
Total	13	33	1 2 3 4	1 17 14 1	3 52 42 3	59	1.55	

Table 4.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Six Reviewers WV 2019 G4B2ELA v2 Number of Assessment Items - 42*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.4 Grade 4 Literary Reading	3	8	16.17	0.75	15.59	8	77.23	13	7.18	7	YES
IR.4 Grade 4 Informational Reading	3	9	14.33	1.21	19.13	11	79.76	9	1.11	3	YES
WL.4 Grade 4 Writing and Language	7	16	28.5	1.87	29.87	15	46.87	19	23.27	14	YES
Total	13	33	59	1.55	23.16	9.8	63.28	12.3	13.56	7.1	

Table 4.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers
WV 2019 G4B2ELA v2 Number of Assessment Items - 42

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.4 Grade 4 Literary Reading	3	8	16.17	0.75	6.67	0.52	83.33	6.45	YES	36	2	0.86	0.03	YES
IR.4 Grade 4 Informational Reading	3	9	14.33	1.21	7.17	0.75	79.63	8.36	YES	32	3	0.78	0.02	YES
WL.4 Grade 4 Writing and Language	7	16	28.5	1.87	6.17	0.98	38.54	6.14	NO	33	3	0.66	0.03	WEAK
Total	13	33	59	1.55	6.7	0.5	67.17	25		34	2	0.77	0.1	

Table 4.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers
WV 2019 G4B2ELA v2 Number of Assessment Items - 42

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.4 Grade 4 Literary Reading	YES	YES	YES	YES
IR.4 Grade 4 Informational Reading	YES	YES	YES	YES
WL.4 Grade 4 Writing and Language	YES	YES	NO	WEAK

Table 4.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV 2019 G4B2ELA v2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6	Reviewer 7
1	1	1	1	1		1	1
2	1	1	1	1		1	1
3	2	1	1	1		1	1
4	2	1	1	1		1	1
5	2	2	2	2		2	2
6	2	2	2	2		2	2
7	3	3	3	3		3	3
8	2	3	2	2		2	2
9	2	2	2	2		2	2
10	3	3	2	2		3	3
11	3	3	2	2		3	2
12	2	2	2	2		2	1
13	2	2	2	2		2	2
14	1	1	2	2		2	1
15	2	2	2	2		2	2
16	3	2	1	2		2	2
17	3	2	3	2		2	2
18	2	1	2	2		2	2
19	2	1	2	2		2	2
20	2	2	2	2		2	2
21	2	1	1	2		2	2
22	2	1	2	2		2	2
23	2	2	2	2		2	2
24	2	1	2	2		2	1
25	2	1	2	2		2	1
26	2	2	2	2		2	2
27	2	2	2	2		2	2
28	2	3	2	2		2	2
29	2	2	2	2		2	2
30	2	2	2	2		2	2

31	2	2	2	2		2	2
32	2	2	2	2		2	2
33	1	1	2	2		2	2
34	3	3	2	2		3	3
35	2	3	2	2		3	3
36	2	1	3	2		3	2
37	1	1	1	1		1	1
38	2	1	1	1		1	1
39	1	1	1	1		1	1
40	2	1	1	2		1	1
41	3	3	3	4		3	3
42	3	3	3	4		3	3

Intraclass correlation - .9385

Pairwise Comparison - 0.74

Table 4.6
DOK Levels and Objectives Code by Each Reviewer
WV 2019 G4B2ELA v2

Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	ELA.4.3 7			1	ELA.4.3 9			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7		
2	1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6		
3	2	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6		
4	2	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6		
5	2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2		
6	2	ELA.4.1			2	ELA.4.1			2	ELA.4.3			2	ELA.4.3			2	ELA.4.1			2	ELA.4.1		
7	3	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4		
8	2	ELA.4.3			3	ELA.4.3			2	ELA.4.3			2	ELA.4.3			2	ELA.4.3			2	ELA.4.7		
9	2	ELA.4.1			2	ELA.4.1			2	ELA.4.3			2	ELA.4.1			2	ELA.4.1			2	ELA.4.3		
10	3	ELA.4.3			3	ELA.4.1 4			2	ELA.4.1 4			2	ELA.4.1 4			3	ELA.4.1 4			3	ELA.4.1 4		
11	3	ELA.4.1 4			3	ELA.4.9			2	ELA.4.9			2	ELA.4.9			3	ELA.4.1 4			2	ELA.4.9		
12	2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			1	ELA.4.9		
13	2	ELA.4.7			2	ELA.4.3			2	ELA.4.7			2	ELA.4.4 0			2	ELA.4.7			2	ELA.4.4 0		
14	1	ELA.4.7			1	ELA.4.7			2	ELA.4.7			2	ELA.4.7			2	ELA.4.7			1	ELA.4.7		
15	2	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5		
16	3	ELA.4.1 1			2	ELA.4.1 1			1	ELA.4.1 1			2	ELA.4.1 1			2	ELA.4.1 1			2	ELA.4.1 1		
17	3	ELA.4.1 5			2	ELA.4.1 5			3	ELA.4.1 5			2	ELA.4.1 5			2	ELA.4.1 5			2	ELA.4.1 5		
18	2	ELA.4.1 0			1	ELA.4.3 9			2	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.1 0		
19	2	ELA.4.1 0			1	ELA.4.3 9			2	ELA.4.1 0			2	ELA.4.3 9			2	ELA.4.1 0			2	ELA.4.1 0		
20	2	ELA.4.4			2	ELA.4.4			2	ELA.4.5			2	ELA.4.4			2	ELA.4.4			2	ELA.4.4		

21	2	ELA.4.4			1	ELA.4.4			1	ELA.4.4			2	ELA.4.4			2	ELA.4.4			2	ELA.4.1 6		
22	2	ELA.4.4			1	ELA.4.4			2	ELA.4.6			2	ELA.4.4			2	ELA.4.4			2	ELA.4.4		
23	2	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5			2	ELA.4.5		
24	2	ELA.4.1 0			1	ELA.4.3 9			2	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.1 0			1	ELA.4.1 0		
25	2	ELA.4.1 0			1	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.1 0			2	ELA.4.1 0			1	ELA.4.3 9		
26	2	ELA.4.6			2	ELA.4.6			2	ELA.4.1 6			2	ELA.4.6			2	ELA.4.4			2	ELA.4.4		
27	2	ELA.4.1 1			2	ELA.4.1 1			2	ELA.4.1 1			2	ELA.4.1 1			2	ELA.4.1 1			2	ELA.4.1 1		
28	2	ELA.4.1 7			3	ELA.4.1 7			2	ELA.4.1 2			2	ELA.4.1 7			2	ELA.4.5			2	ELA.4.5		
29	2	ELA.4.4			2	ELA.4.6			2	ELA.4.6			2	ELA.4.1 6			2	ELA.4.6			2	ELA.4.6		
30	2	ELA.4.7			2	ELA.4.4 0			2	ELA.4.3 9			2	ELA.4.7			2	ELA.4.7			2	ELA.4.7		
31	2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2			2	ELA.4.2		
32	2	ELA.4.3			2	ELA.4.3			2	ELA.4.1			2	ELA.4.1			2	ELA.4.1			2	ELA.4.4		
33	1	ELA.4.7			1	ELA.4.7			2	ELA.4.7			2	ELA.4.7			2	ELA.4.7			2	ELA.4.7		
34	3	ELA.4.3			3	ELA.4.3			2	ELA.4.3			2	ELA.4.3			3	ELA.4.3			3	ELA.4.1		
35	2	ELA.4.1 3			3	ELA.4.1 3			2	ELA.4.1 3			2	ELA.4.1 3			3	ELA.4.1 3			3	ELA.4.1 3		
36	2	ELA.4.1 3			1	ELA.4.1 3			3	ELA.4.1 3			2	ELA.4.1 3			3	ELA.4.1 3			2	ELA.4.1 3		
37	1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7			1	ELA.4.3 7		
38	2	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 7		
39	1	ELA.4.3 7			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 6			1	ELA.4.3 7			1	ELA.4.3 7		
40	2	ELA.4.3 6	ELA.4.3 7		1	ELA.4.3 7	ELA.4.3 6		1	ELA.4.3 6	ELA.4.3 7		2	ELA.4.3 6	ELA.4.3 7		1	ELA.4.3 6	ELA.4.3 7		1	ELA.4.3 6	ELA.4.3 7	
41	3	ELA.4.2 1	ELA.4.3 8		3	ELA.4.3 8			3	ELA.4.2 1	ELA.4.3 8		4	ELA.4.2 1	ELA.4.3 8		3	ELA.4.2 1	ELA.4.3 8		3	ELA.4.2 1	ELA.4.3 6	
42	3	ELA.4.2 1	ELA.4.2 3	ELA.4.3 8	3	ELA.4.2 1	ELA.4.3 8	ELA.4.2 3	3	ELA.4.2 1	ELA.4.2 3		4	ELA.4.2 1	ELA.4.2 3		3	ELA.4.2 1	ELA.4.2 3		3	ELA.4.3 8	ELA.4.2 1	ELA.4.2 3
Objective Pairwise Comparison: 0.7																								
Standard Pairwise Comparison: 0.93																								

Table 4.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV 2019 G4B2ELA v2

	Low		Medium		High		
	0		10.8			18	
LR.4							
LR.4.KID							
ELA.4.1	6(4)	9(4)	34(1)	32(3)			
ELA.4.2	5(6)	12(5)	31(6)				
ELA.4.3	13(1)	6(2)	9(2)	10(1)	8(5)	32(2)	34(5)
LR.4.CS							
ELA.4.7	33(6)	8(1)	14(6)	13(3)	30(4)		
ELA.4.8							
ELA.4.9	12(1)	11(4)					
LR.4.IKI							
ELA.4.13	35(6)	36(6)					
ELA.4.14	11(2)	10(5)	7(6)				
IR.4							
IR.4.KID							
ELA.4.4	20(5)	21(5)	29(1)	22(5)	26(2)	32(1)	
ELA.4.5	28(2)	20(1)	23(6)	15(6)			
ELA.4.6	29(4)	26(3)	22(1)				
IR.4.CS							
ELA.4.10	25(5)	24(5)	18(5)	19(4)			
ELA.4.11	27(6)	16(6)					
ELA.4.12	28(1)						
IR.4.IKI							
ELA.4.15	17(6)						
ELA.4.16	29(1)	26(1)	21(1)				

ELA.4.17	28(3)						
WL.4							
WL.4.TTP							
ELA.4.20							
ELA.4.21	41(15)	42(18)					
ELA.4.22							
WL.4.PDW							
ELA.4.23	42(18)						
ELA.4.24							
ELA.4.25							
WL.4.RBPK							
ELA.4.26							
ELA.4.27							
ELA.4.28							
WL.4.RW							
ELA.4.29							
WL.4.CSE							
ELA.4.36	38(10)	39(6)	2(6)	3(6)	4(12)	41(3)	40(12)
ELA.4.37	1(5)	39(6)	38(2)	37(6)	40(12)		
WL.4.KL							
ELA.4.38	41(15)	42(9)					
WL.4.VAU							
ELA.4.39	1(1)	30(1)	25(1)	24(1)	19(2)	18(1)	
ELA.4.40	13(2)	30(1)					
ELA.4.41							

Table 4.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 WV 2019 G4B2ELA v2

Low		Medium		High	
3.6		10.8		18	
1	10199-3064	ELA.4.37:5	ELA.4.39:1		
2	10199-3071	ELA.4.36:6			
3	10199-3073	ELA.4.36:6			
4	10199-3075	ELA.4.36:12			
5	10199-7486	ELA.4.2:6			
6	10199-7487	ELA.4.1:4	ELA.4.3:2		
7	10199-7489	ELA.4.14:6			
8	10199-7490	ELA.4.3:5	ELA.4.7:1		
9	10199-7491	ELA.4.1:4	ELA.4.3:2		
10	10199-7492	ELA.4.3:1	ELA.4.14:5		
11	10199-7493	ELA.4.9:4	ELA.4.14:2		
12	10199-7494	ELA.4.2:5	ELA.4.9:1		
13	10199-7495	ELA.4.3:1	ELA.4.7:3	ELA.4.40:2	
14	10199-7496	ELA.4.7:6			
15	REP10199-13913	ELA.4.5:6			
16	REP10199-13921	ELA.4.11:6			
17	REP10199-13923	ELA.4.15:6			
18	REP10199-13928	ELA.4.10:5	ELA.4.39:1		
19	REP10199-13929	ELA.4.10:4	ELA.4.39:2		
20	REP10199-13934	ELA.4.4:5	ELA.4.5:1		
21	10199-7578	ELA.4.4:5	ELA.4.16:1		
22	10199-7579	ELA.4.4:5	ELA.4.6:1		
23	10199-7580	ELA.4.5:6			
24	10199-7581	ELA.4.10:5	ELA.4.39:1		
25	10199-7583	ELA.4.10:5	ELA.4.39:1		
26	10199-7584	ELA.4.4:2	ELA.4.6:3	ELA.4.16:1	
27	10199-7587	ELA.4.11:6			
28	10199-7599	ELA.4.5:2	ELA.4.12:1	ELA.4.17:3	
29	10199-7600	ELA.4.4:1	ELA.4.6:4	ELA.4.16:1	
30	REP10199-3350	ELA.4.7:4	ELA.4.39:1	ELA.4.40:1	

31 REP10199-3479	ELA.4.2:6		
32 10199-3481	ELA.4.1:3	ELA.4.3:2	ELA.4.4:1
33 REP10199-3667	ELA.4.7:6		
34 REP10199-3686	ELA.4.1:1	ELA.4.3:5	
35 REP10199-7065	ELA.4.13:6		
36 REP10199-7066	ELA.4.13:6		
37 10199-3091	ELA.4.37:6		
38 10199-3092	ELA.4.36:10	ELA.4.37:2	
39 10199-3093	ELA.4.36:6	ELA.4.37:6	
40 CONVENTIONS10199-14163	ELA.4.36:12	ELA.4.37:12	
41 ELABORATION10199-14163	ELA.4.21:15	ELA.4.36:3	ELA.4.38:15
42 ORGANIZATION10199-14163	ELA.4.21:18	ELA.4.23:18	ELA.4.38:9

Table 4.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV 2019 G4B2ELA v2

Low DOK		Matched DOK		High DOK

LR.4							
LR.4.KID							
ELA.4.1: [2]	6:(4)[2]	9:(4)[2]	32:(3)[2]	34:(1)[3]			
ELA.4.2: [2]	5:(6)[2]	12:(5)[2]	31:(6)[2]				
ELA.4.3: [2]	6:(2)[2]	8:(5)[2]	9:(2)[2]	10:(1)[3]	13:(1)[2]	32:(2)[2]	34:(5)[3]
LR.4.CS							
ELA.4.7: [2]	8:(1)[2]	13:(3)[2]	14:(6)[2]	30:(4)[2]	33:(6)[2]		
ELA.4.8							
ELA.4.9: [2]	11:(4)[2]	12:(1)[1]					
LR.4.IKI							
ELA.4.13: [3]	35:(6)[2]	36:(6)[2]					
ELA.4.14: [3]	7:(6)[3]	10:(5)[3]	11:(2)[3]				
IR.4							
IR.4.KID							
ELA.4.4: [2]	20:(5)[2]	21:(5)[2]	22:(5)[2]	26:(2)[2]	29:(1)[2]	32:(1)[2]	
ELA.4.5: [2]	15:(6)[2]	20:(1)[2]	23:(6)[2]	28:(2)[2]			
ELA.4.6: [2]	22:(1)[2]	26:(3)[2]	29:(4)[2]				
IR.4.CS							
ELA.4.10: [2]	18:(5)[2]	19:(4)[2]	24:(5)[2]	25:(5)[2]			
ELA.4.11: [2]	16:(6)[2]	27:(6)[2]					
ELA.4.12: [3]	28:(1)[2]						
IR.4.IKI							
ELA.4.15: [3]	17:(6)[2]						
ELA.4.16: [3]	21:(1)[2]	26:(1)[2]	29:(1)[2]				
ELA.4.17: [3]	28:(3)[2]						
WL.4							
WL.4.TTP							
ELA.4.20							

ELA.4.21: [3]	41:(15)[3]	42:(18)[3]					
ELA.4.22							
WL.4.PDW							
ELA.4.23: [3]	42:(18)[3]						
ELA.4.24							
ELA.4.25							
WL.4.RBPK							
ELA.4.26							
ELA.4.27							
ELA.4.28							
WL.4.RW							
ELA.4.29							
WL.4.CSE							
ELA.4.36: [2]	2:(6)[1]	3:(6)[1]	4:(12)[1]	38:(10)[1]	39:(6)[1]	40:(12)[1]	41:(3)[3]
ELA.4.37: [1]	1:(5)[1]	37:(6)[1]	38:(2)[1]	39:(6)[1]	40:(12)[1]		
WL.4.KL							
ELA.4.38: [2]	41:(15)[3]	42:(9)[3]					
WL.4.VAU							
ELA.4.39: [2]	1:(1)[1]	18:(1)[1]	19:(2)[2]	24:(1)[1]	25:(1)[1]	30:(1)[2]	
ELA.4.40: [3]	13:(2)[2]	30:(1)[2]					
ELA.4.41							

Table 4.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
CONF WV Grade 4B2 ELA

Low	Medium	High
3	9	15

1 10199-3064	Exact:6	
2 10199-3071	Exact:6	
3 10199-3073	Exact:6	
4 10199-3075	Exact:12	
5 10199-7486	Exact:6	
6 10199-7487	Exact:6	
7 10199-7489	Exact:6	
8 10199-7490	Exact:6	
9 10199-7491	Exact:5	Partial:1
10 10199-7492	Exact:6	
11 10199-7493	Exact:6	
12 10199-7494	Exact:6	
13 10199-7495	Exact:6	
14 10199-7496	Exact:6	
15 REP10199-13913	Exact:6	
16 REP10199-13921	Exact:6	
17 REP10199-13923	Exact:6	
18 REP10199-13928	Exact:6	
19 REP10199-13929	Exact:5	Partial:1
20 REP10199-13934	Exact:6	
21 10199-7578	Exact:6	
22 10199-7579	Exact:6	
23 10199-7580	Exact:6	
24 10199-7581	Exact:6	
25 10199-7583	Exact:6	
26 10199-7584	Exact:6	
27 10199-7587	Exact:6	
28 10199-7599	Exact:6	
29 10199-7600	Exact:6	
30 REP10199-3350	Exact:6	

31 REP10199-3479	Exact:6	
32 10199-3481	Exact:6	
33 REP10199-3667	Exact:6	
34 REP10199-3686	Exact:6	
35 REP10199-7065	Exact:6	
36 REP10199-7066	Exact:6	
37 10199-3091	Exact:6	
38 10199-3092	Exact:12	
39 10199-3093	Exact:12	
40 CONVENTIONS10199-14163	Partial:2	Minimal:10
41 ELABORATION10199-14163	Partial:15	Minimal:3
42 ORGANIZATION10199-14163	Partial:15	Minimal:3

Grade 5 Batch 1 ELA West Virginia

Table 5.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 5 B1*
Number of Assessment Items - 44

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.5 Grade 5 Literary Reading	3	8	2 3	4 4	50 50	17.33	1.37	YES
IR.5 Grade 5 Informational Reading	3	9	2 3	5 4	55.56 44.44	15.17	0.41	YES
WL.5 Grade 5 Writing and Language	7	16	1 2 3 4	1 7 7 1	6.25 43.75 43.75 6.25	27.33	2.07	YES
Total	13	33	1 2 3 4	1 16 15 1	3 48 45 3	59.83	2.32	

Table 5.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 5 B1*
Number of Assessment Items - 44

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.5 Grade 5 Literary Reading	3	8	17.33	1.37	15.82	13	80.29	14	3.89	5	YES
IR.5 Grade 5 Informational Reading	3	9	15.17	0.41	7.57	10	87.99	8	4.44	7	YES
WL.5 Grade 5 Writing and Language	7	16	27.33	2.07	18.35	9	59.23	14	22.42	16	YES
Total	13	33	59.83	2.32	15.04	6.3	72.42	9.9	12.53	6.9	

Table 5.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers
WV ELA 2019 Grade 5 B1 Number of Assessment Items - 44

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.5 Grade 5 Literary Reading	3	8	17.33	1.37	6.17	0.75	77.08	9.41	YES	36	2	0.81	0.06	YES
IR.5 Grade 5 Informational Reading	3	9	15.17	0.41	7.83	0.75	87.04	8.36	YES	32	1	0.79	0.07	YES
WL.5 Grade 5 Writing and Language	7	16	27.33	2.07	5.67	0.52	35.42	3.23	NO	32	2	0.76	0.08	YES
Total	13	33	59.83	2.32	6.6	1.13	66.51	27		33	3	0.79	0.03	

Table 5.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers
WV ELA 2019 Grade 5 B1 Number of Assessment Items - 44

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.5 Grade 5 Literary Reading	YES	YES	YES	YES
IR.5 Grade 5 Informational Reading	YES	YES	YES	YES
WL.5 Grade 5 Writing and Language	YES	YES	NO	YES

Table 5.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation WV ELA 2019 Grade 5 B1*

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	2	2	2	2	2
2	2	2	2	2	2	2
3	2	3	2	2	2	2
4	2	2	2	2	2	2
5	2	2	2	3	2	3
6	2	3	2	2	3	2
7	3	2	2	3	3	3
8	2	2	2	2	2	2
9	2	3	2	2	2	2
10	2	2	2	2	2	2
11	3	2	2	2	2	2
12	3	2	2	3	3	3
13	2	2	2	2	2	2
14	3	3	2	3	3	2
15	2	2	3	2	2	2
16	2	3	2	3	3	3
17	1	1	2	1	1	2
18	2	1	1	1	1	2
19	1	1	1	1	1	1
20	1	1	1	1	1	1
21	2	1	2	2	2	2
22	2	3	2	3	2	3
23	3	2	2	2	2	2
24	3	3	2	3	3	3
25	2	2	2	2	2	3
26	2	2	2	2	2	2
27	3	2	2	3	3	3
28	2	3	3	3	2	3
29	3	2	2	2	2	2
30	2	2	2	2	2	2
31	2	2	2	2	2	2

32	2	3	2	2	2	2
33	2	2	2	2	2	2
34	3	2	2	3	3	3
35	2	2	1	2	2	2
36	2	2	2	3	2	2
37	2	2	2	2	2	2
38	2	1	1	1	1	2
39	1	1	1	1	1	1
40	2	1	1	1	1	1
41	2	1	1	1	1	1
42	2	1	2	1	1	1
43	3	3	4	3	3	3
44	3	3	4	3	3	3

Intraclass correlation - .9356

Pairwise Comparison - 0.71

Table 5.6 DOK Levels and Objectives Code by Each Reviewer WV ELA 2019 Grade 5 B1 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.5.1			2	ELA.5.4			2	ELA.5.4			2	ELA.5.4			2	ELA.5.6			2	ELA.5.4		
2	2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5		
3	2	ELA.5.10			3	ELA.5.40			2	ELA.5.40			2	ELA.5.10			2	ELA.5.10			2	ELA.5.39		
4	2	ELA.5.10			2	ELA.5.10			2	ELA.5.10			2	ELA.5.10			2	ELA.5.10			2	ELA.5.10		
5	2	ELA.5.4			2	ELA.5.4			2	ELA.5.4			3	ELA.5.16			2	ELA.5.16			3	ELA.5.16		
6	2	ELA.5.4			3	ELA.5.16			2	ELA.5.16			2	ELA.5.6			3	ELA.5.16			2	ELA.5.5		
7	3	ELA.5.3			2	ELA.5.3			2	ELA.5.14			3	ELA.5.3			3	ELA.5.3			3	ELA.5.14		
8	2	ELA.5.8			2	ELA.5.2			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8		
9	2	ELA.5.2			3	ELA.5.1			2	ELA.5.1			2	ELA.5.1			2	ELA.5.1			2	ELA.5.1		
10	2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8		
11	3	ELA.5.9			2	ELA.5.9			2	ELA.5.9			2	ELA.5.9			2	ELA.5.9			2	ELA.5.9		
12	3	ELA.5.2			2	ELA.5.1			2	ELA.5.2	ELA.5.1		3	ELA.5.2			3	ELA.5.2			3	ELA.5.3		
13	2	ELA.5.7			2	ELA.5.40			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7		
14	3	ELA.5.3			3	ELA.5.3			2	ELA.5.2			3	ELA.5.3			3	ELA.5.2			2	ELA.5.1		
15	2	ELA.5.9			2	ELA.5.1			3	ELA.5.14			2	ELA.5.1			2	ELA.5.1			2	ELA.5.8		
16	2	ELA.5.3			3	ELA.5.3			2	ELA.5.1			3	ELA.5.3			3	ELA.5.3			3	ELA.5.3		
17	1	ELA.5.36			1	ELA.5.36			2	ELA.5.36			1	ELA.5.36			1	ELA.5.36			2	ELA.5.36		
18	2	ELA.5.36			1	ELA.5.36			1	ELA.5.36			1	ELA.5.36			1	ELA.5.36			2	ELA.5.36		
19	1	ELA.5.37			1	ELA.5.37			1	ELA.5.37			1	ELA.5.37			1	ELA.5.37			1	ELA.5.37		
20	1	ELA.5.37			1	ELA.5.37			1	ELA.5.37			1	ELA.5.37			1	ELA.5.37			1	ELA.5.37		
21	2	ELA.5.5			1	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5		
22	2	ELA.5.4			3	ELA.5.16			2	ELA.5.5			3	ELA.5.16			2	ELA.5.16			3	ELA.5.16		

23	3	ELA.5.6			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			
24	3	ELA.5.1 2		3	ELA.5.1 2		2	ELA.5.1 2		3	ELA.5.1 2		3	ELA.5.1 2		3	ELA.5.1 2				
25	2	ELA.5.4			2	ELA.5.6			2	ELA.5.6			2	ELA.5.4			2	ELA.5.4		3	ELA.5.1 6
26	2	ELA.5.1 1			2	ELA.5.1 1			2	ELA.5.1 1			2	ELA.5.1 1			2	ELA.5.1 1		2	ELA.5.1 1
27	3	ELA.5.1 7			2	ELA.5.4			2	ELA.5.1 5			3	ELA.5.1 2			3	ELA.5.1 5		3	ELA.5.5
28	2	ELA.5.4			3	ELA.5.1 7			3	ELA.5.1 7			3	ELA.5.1 7			2	ELA.5.1 5		3	ELA.5.5
29	3	ELA.5.6			2	ELA.5.6			2	ELA.5.6			2	ELA.5.6			2	ELA.5.6		2	ELA.5.6
30	2	ELA.5.1 0			2	ELA.5.1 0			2	ELA.5.1 0			2	ELA.5.3 9			2	ELA.5.1 0		2	ELA.5.1 0
31	2	ELA.5.1			2	ELA.5.1			2	ELA.5.3			2	ELA.5.2			2	ELA.5.2		2	ELA.5.1
32	2	ELA.5.3		3	ELA.5.4 0		2	ELA.5.7		2	ELA.5.7		2	ELA.5.7		2	ELA.5.7		2	ELA.5.7	
33	2	ELA.5.1			2	ELA.5.1			2	ELA.5.3			2	ELA.5.8			2	ELA.5.9		2	ELA.5.9
34	3	ELA.5.3		2	ELA.5.8		2	ELA.5.8		3	ELA.5.8		3	ELA.5.8		3	ELA.5.2		3	ELA.5.2	
35	2	ELA.5.1			2	ELA.5.1			1	ELA.5.1			2	ELA.5.2			2	ELA.5.9		2	ELA.5.1
36	2	ELA.5.9			2	ELA.5.9			2	ELA.5.9			3	ELA.5.9			2	ELA.5.9		2	ELA.5.9
37	2	ELA.5.3			2	ELA.5.1			2	ELA.5.1			2	ELA.5.1			2	ELA.5.1		2	ELA.5.1
38	2	ELA.5.3 6		1	ELA.5.3 6		1	ELA.5.3 6		1	ELA.5.3 6		1	ELA.5.3 6			1	ELA.5.3 6		2	ELA.5.3 6
39	1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7		1	ELA.5.3 7
40	2	ELA.5.3 6		1	ELA.5.3 6		1	ELA.5.3 7		1	ELA.5.3 6		1	ELA.5.3 6		1	ELA.5.3 6		1	ELA.5.3 6	
41	2	ELA.5.3 6		1	ELA.5.3 6		1	ELA.5.3 6		1	ELA.5.3			1	ELA.5.3 6		1	ELA.5.3 6		1	ELA.5.3 6
42	2	ELA.5.3 6	ELA.5.3 7		1	ELA.5.3 6	ELA.5.3 7		2	ELA.5.3 6	ELA.5.3 7		1	ELA.5.3 6	ELA.5.3 7		1	ELA.5.3 6	ELA.5.3 7	1	ELA.5.3 7
43	3	ELA.5.2 0	ELA.5.3 8		3	ELA.5.2 0	ELA.5.3 8		4	ELA.5.2 0	ELA.5.3 8		3	ELA.5.2 1	ELA.5.3 8		3	ELA.5.2 0	ELA.5.3 8	3	ELA.5.3 8
44	3	ELA.5.3 8	ELA.5.2 0	ELA.5.2 3	3	ELA.5.2 0	ELA.5.2 3		4	ELA.5.2 0	ELA.5.2 3		3	ELA.5.3 8	ELA.5.2 3	ELA.5.2 1	3	ELA.5.2 0	ELA.5.2 3	3	ELA.5.2 1
Objective Pairwise Comparison: 0.61																					
Standard Pairwise Comparison: 0.94																					

Table 5.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers) WV ELA 2019 Grade 5 B1

Low	Medium			High
0	10.8			18

LR.5											
LR.5.KID											
ELA.5.1	1(1)	9(5)	14(1)	15(3)	16(1)	31(3)	33(2)	35(4)	37(5)	12(2)	
ELA.5.2	35(1)	34(2)	31(2)	14(2)	12(4)	9(1)	8(1)				
ELA.5.3	7(4)	12(1)	14(3)	31(1)	32(1)	16(5)	17(1)	34(1)	33(1)	37(1)	41(1)
LR.5.CS											
ELA.5.7	32(4)	13(5)									
ELA.5.8	15(1)	10(6)	8(5)	33(1)	34(3)						
ELA.5.9	33(2)	35(1)	36(6)	11(6)	15(1)						
LR.5.IKI											
ELA.5.13											
ELA.5.14	15(1)	7(2)									
IR.5											
IR.5.KID											
ELA.5.4	5(3)	6(1)	1(4)	25(3)	27(1)	28(1)	22(1)				
ELA.5.5	22(1)	21(6)	23(5)	28(1)	27(1)	2(6)	6(1)				
ELA.5.6	6(1)	1(1)	25(2)	29(6)	23(1)						
IR.5.CS											
ELA.5.10	30(5)	4(6)	3(3)								
ELA.5.11	26(6)										
ELA.5.12	27(1)	24(6)									
IR.5.IKI											

ELA.5.15	27(2)	28(1)									
ELA.5.16	25(1)	22(4)	6(3)	5(3)							
ELA.5.17	28(3)	27(1)									
WL.5											
WL.5.TTP											
ELA.5.20	43(12)	44(12)									
ELA.5.21	43(3)	44(6)									
ELA.5.22											
WL.5.PDW											
ELA.5.23	44(15)										
ELA.5.24											
ELA.5.25											
WL.5.RBPK											
ELA.5.26											
ELA.5.27											
ELA.5.28	44(3)										
WL.5.RW											
ELA.5.29											
WL.5.CSE											
ELA.5.36	40(5)	41(5)	38(6)	17(6)	18(6)	42(12)					
ELA.5.37	19(12)	20(6)	39(12)	40(1)	42(12)						
WL.5.KL											
ELA.5.38	44(6)	43(18)									
WL.5.VAU											
ELA.5.39	30(1)	3(1)									
ELA.5.40	3(2)	13(1)	32(1)								
ELA.5.41											

Table 5.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

WV ELA 2019 Grade 5 B1

Low		Medium		High	
3.6		10.8		18	
1	10199-3423	ELA.5.1:1	ELA.5.4:4	ELA.5.6:1	
2	10199-3441	ELA.5.5:6			
3	10199-3443	ELA.5.10:3	ELA.5.39:1	ELA.5.40:2	
4	10199-3604	ELA.5.10:6			
5	10199-3615	ELA.5.4:3	ELA.5.16:3		
6	10199-6705	ELA.5.4:1	ELA.5.5:1	ELA.5.6:1	ELA.5.16:3
7	10199-5600	ELA.5.3:4	ELA.5.14:2		
8	10199-5603	ELA.5.2:1	ELA.5.8:5		
9	10199-5604	ELA.5.1:5	ELA.5.2:1		
10	10199-5605	ELA.5.8:6			
11	10199-5606	ELA.5.9:6			
12	10199-5607	ELA.5.1:2	ELA.5.2:4	ELA.5.3:1	
13	10199-5609	ELA.5.7:5	ELA.5.40:1		
14	10199-5610	ELA.5.1:1	ELA.5.2:2	ELA.5.3:3	
15	10199-5612	ELA.5.1:3	ELA.5.8:1	ELA.5.9:1	ELA.5.14:1
16	10199-5734	ELA.5.1:1	ELA.5.3:5		
17	10199-3119	ELA.5.3:1	ELA.5.36:6		
18	10199-3120	ELA.5.36:6			
19	10199-3121	ELA.5.37:12			
20	10199-3122	ELA.5.37:6			
21	10199-8759	ELA.5.5:6			
22	REP10199-8762	ELA.5.4:1	ELA.5.5:1	ELA.5.16:4	
23	REP10199-10890	ELA.5.5:5	ELA.5.6:1		
24	REP10199-10923	ELA.5.12:6			
25	REP10199-11106	ELA.5.4:3	ELA.5.6:2	ELA.5.16:1	
26	REP10199-11107	ELA.5.11:6			

27 REP10199-11109	ELA.5.4:1	ELA.5.5:1	ELA.5.12:1	ELA.5.15:2	ELA.5.17:1
28 REP10199-11110	ELA.5.4:1	ELA.5.5:1	ELA.5.15:1	ELA.5.17:3	
29 REP10199-13973	ELA.5.6:6				
30 REP10199-13979	ELA.5.10:5	ELA.5.39:1			
31 REP10199-3401	ELA.5.1:3	ELA.5.2:2	ELA.5.3:1		
32 REP10199-3402	ELA.5.3:1	ELA.5.7:4	ELA.5.40:1		
33 REP10199-3425	ELA.5.1:2	ELA.5.3:1	ELA.5.8:1	ELA.5.9:2	
34 10199-3404	ELA.5.2:2	ELA.5.3:1	ELA.5.8:3		
35 REP10199-3416	ELA.5.1:4	ELA.5.2:1	ELA.5.9:1		
36 REP10199-4839	ELA.5.9:6				
37 REP10199-3732	ELA.5.1:5	ELA.5.3:1			
38 10199-1350	ELA.5.36:6				
39 10199-1351	ELA.5.37:1 2				
40 10199-1353	ELA.5.36:5	ELA.5.37:1			
41 10199-1354	ELA.5.3:1	ELA.5.36:5			
42 CONVENTIONS10199-14172	ELA.5.36:1 2	ELA.5.37:1 2			
43 ELABORATION10199-14172	ELA.5.20:1 2	ELA.5.21:3	ELA.5.38:1 8		
44 ORGANIZATION10199-14172	ELA.5.20:1 2	ELA.5.21:6	ELA.5.23:1 5	ELA.5.28:3	ELA.5.38:6

Table 5.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 5 B1

Low DOK		Matched DOK		High DOK

LR.5											
LR.5.KID											
ELA.5.1: [2]	1:(1)[2]	9:(5)[2]	12:(2)[2]	14:(1)[2]	15:(3)[2]	16:(1)[2]	31:(3)[2]	33:(2)[2]	35:(4)[2]	37:(5)[2]	
ELA.5.2: [3]	8:(1)[2]	9:(1)[2]	12:(4)[3]	14:(2)[2]	31:(2)[2]	34:(2)[3]	35:(1)[2]				
ELA.5.3: [3]	7:(4)[3]	12:(1)[3]	14:(3)[3]	16:(5)[3]	17:(1)[2]	31:(1)[2]	32:(1)[2]	33:(1)[2]	34:(1)[3]	37:(1)[2]	41:(1)[1]
LR.5.CS											
ELA.5.7: [2]	13:(5)[2]	32:(4)[2]									
ELA.5.8: [2]	8:(5)[2]	10:(6)[2]	15:(1)[2]	33:(1)[2]	34:(3)[2]						
ELA.5.9: [2]	11:(6)[2]	15:(1)[2]	33:(2)[2]	35:(1)[2]	36:(6)[2]						
LR.5.IKI											
ELA.5.13											
ELA.5.14: [3]	7:(2)[2]	15:(1)[3]									
IR.5											
IR.5.KID											
ELA.5.4: [2]	1:(4)[2]	5:(3)[2]	6:(1)[2]	22:(1)[2]	25:(3)[2]	27:(1)[2]	28:(1)[2]				
ELA.5.5: [2]	2:(6)[2]	6:(1)[2]	21:(6)[2]	22:(1)[2]	23:(5)[2]	27:(1)[3]	28:(1)[3]				
ELA.5.6: [2]	1:(1)[2]	6:(1)[2]	23:(1)[3]	25:(2)[2]	29:(6)[2]						
IR.5.CS											
ELA.5.10: [2]	3:(3)[2]	4:(6)[2]	30:(5)[2]								
ELA.5.11: [2]	26:(6)[2]										
ELA.5.12: [3]	24:(6)[3]	27:(1)[3]									
IR.5.IKI											
ELA.5.15: [3]	27:(2)[2]	28:(1)[2]									
ELA.5.16: [3]	5:(3)[3]	6:(3)[3]	22:(4)[3]	25:(1)[3]							
ELA.5.17: [3]	27:(1)[3]	28:(3)[3]									

WL.5											
WL.5.TTP											
ELA.5.20: [3]	43:(12)[3]	44:(12)[3]									
ELA.5.21: [3]	43:(3)[3]	44:(6)[3]									
ELA.5.22											
WL.5.PDW											
ELA.5.23: [3]	44:(15)[3]										
ELA.5.24											
ELA.5.25											
WL.5.RBPK											
ELA.5.26											
ELA.5.27											
ELA.5.28: [3]	44:(3)[3]										
WL.5.RW											
ELA.5.29											
WL.5.CSE											
ELA.5.36: [2]	17:(6)[1]	18:(6)[1]	38:(6)[1]	40:(5)[1]	41:(5)[1]	42:(12)[1]					
ELA.5.37: [1]	19:(12)[1]	20:(6)[1]	39:(12)[1]	40:(1)[1]	42:(12)[1]						
WL.5.KL											
ELA.5.38: [2]	43:(18)[3]	44:(6)[3]									
WL.5.VAU											
ELA.5.39: [2]	3:(1)[2]	30:(1)[2]									
ELA.5.40: [3]	3:(2)[2]	13:(1)[2]	32:(1)[3]								
ELA.5.41											

Grade 5 Batch 2 ELA West Virginia

Table 5.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 5 B2 Number of Assessment Items - 42*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.5 Grade 5 Literary Reading	3	8	2 3	4 4	50 50	17.33	0.52	YES
IR.5 Grade 5 Informational Reading	3	9	2 3	5 4	55.56 44.44	13.33	0.82	YES
WL.5 Grade 5 Writing and Language	7	16	1 2 3 4	1 7 7 1	6.25 43.75 43.75 6.25	27.17	2.23	YES
Total	13	33	1 2 3 4	1 16 15 1	3 48 45 3	57.83	2.32	

Table 5.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 5 B2 Number of Assessment Items - 42*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.5 Grade 5 Literary Reading	3	8	17.33	0.52	13.56	12	78.76	14	7.68	12	YES
IR.5 Grade 5 Informational Reading	3	9	13.33	0.82	8.61	13	83.68	13	7.71	8	YES
WL.5 Grade 5 Writing and Language	7	16	27.17	2.23	22.18	15	53.98	13	23.85	17	YES
Total	13	33	57.83	2.32	16.14	6.4	68.59	9.4	15.27	8.5	

Table 5.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers
WV ELA 2019 Grade 5 B2 Number of Assessment Items - 42

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.5 Grade 5 Literary Reading	3	8	17.33	0.52	6.33	0.52	79.17	6.45	YES	38	2	0.8	0.06	YES
IR.5 Grade 5 Informational Reading	3	9	13.33	0.82	6.67	0.82	74.07	9.07	YES	29	2	0.85	0.06	YES
WL.5 Grade 5 Writing and Language	7	16	27.17	2.23	5.33	0.52	33.33	3.23	NO	33	2	0.69	0.04	WEAK
Total	13	33	57.83	2.32	6.1	0.69	62.19	25		33	4	0.78	0.08	

Table 5.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers
WV ELA 2019 Grade 5 B2 Number of Assessment Items - 42

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.5 Grade 5 Literary Reading	YES	YES	YES	YES
IR.5 Grade 5 Informational Reading	YES	YES	YES	YES
WL.5 Grade 5 Writing and Language	YES	YES	NO	WEAK

Table 5.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 5 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	1	1	1	1	1
2	2	1	1	1	1	1
3	2	1	1	1	1	1
4	1	1	1	1	1	1
5	3	3	2	2	2	3
6	2	2	2	2	2	2
7	2	2	2	1	2	2
8	2	3	2	2	2	2
9	3	2	2	3	2	3
10	3	3	2	2	3	2
11	2	2	2	2	2	2
12	3	3	2	2	2	3
13	2	2	2	1	2	2
14	2	2	2	3	2	2
15	2	2	2	2	2	2
16	2	3	2	3	3	3
17	2	2	2	2	2	2
18	2	2	2	2	2	2
19	2	2	2	2	2	2
20	2	3	2	2	2	3
21	2	3	2	3	2	3
22	2	2	2	2	2	2
23	2	2	2	2	2	3
24	2	2	2	2	2	3
25	2	2	2	3	2	2
26	2	2	2	3	2	3
27	2	3	2	3	2	3
28	2	2	2	2	2	2
29	2	2	2	3	2	3
30	3	2	2	3	2	3
31	2	2	2	1	2	2

32	2	2	2	3	2	2
33	2	2	2	3	3	3
34	2	2	2	2	2	2
35	2	2	2	2	2	2
36	1	1	1	1	1	1
37	2	1	1	1	1	1
38	2	1	1	1	1	1
39	1	1	1	1	1	1
40	2	1	2	1	1	1
41	3	3	4	3	3	3
42	3	3	4	3	3	3

Intraclass correlation - .9341

Pairwise Comparison - 0.7

Table 5.6
DOK Levels and Objectives Code by Each Reviewer
WV ELA 2019 Grade 5 B2

Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 7			1	ELA.5.3 6			1	ELA.5.3 6		
2	2	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 6		
3	2	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 6		
4	1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 6			1	ELA.5.3 7			1	ELA.5.3 6			1	ELA.5.3 6		
5	3	ELA.5.1 1			3	ELA.5.1 1			2	ELA.5.6			2	ELA.5.1 1			2	ELA.5.1 1			3	ELA.5.1 1		
6	2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5		
7	2	ELA.5.1 0			2	ELA.5.1 0			2	ELA.5.1 0			1	ELA.5.1 0			2	ELA.5.1 0			2	ELA.5.1 0		
8	2	ELA.5.1 1			3	ELA.5.1 2			2	ELA.5.1 2			2	ELA.5.4			2	ELA.5.1 1			2	ELA.5.1 1		
9	3	ELA.5.1 1			2	ELA.5.1 1			2	ELA.5.1 1			3	ELA.5.1 1			2	ELA.5.1 1			3	ELA.5.1 7		
10	3	ELA.5.1 6			3	ELA.5.1 6			2	ELA.5.1 6			2	ELA.5.1 6			3	ELA.5.1 6			2	ELA.5.6		
11	2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5			2	ELA.5.5		
12	3	ELA.5.4			3	ELA.5.1 2			2	ELA.5.1 7			2	ELA.5.6			2	ELA.5.1 1			3	ELA.5.1 7		
13	2	ELA.5.1 0			2	ELA.5.1 0			2	ELA.5.1 0			1	ELA.5.1 0			2	ELA.5.1 0			2	ELA.5.3 9		
14	2	ELA.5.1			2	ELA.5.1			2	ELA.5.1			3	ELA.5.9			2	ELA.5.1			2	ELA.5.1		
15	2	ELA.5.2			2	ELA.5.1			2	ELA.5.1			2	ELA.5.1			2	ELA.5.1			2	ELA.5.7		
16	2	ELA.5.2			3	ELA.5.2			2	ELA.5.2			3	ELA.5.2			3	ELA.5.2			3	ELA.5.2		
17	2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7		
18	2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8		
19	2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8			2	ELA.5.8		
20	2	ELA.5.9			3	ELA.5.3			2	ELA.5.3			2	ELA.5.3			2	ELA.5.3			3	ELA.5.3		

21	2	ELA.5.1 4			3	ELA.5.2			2	ELA.5.1 4			3	ELA.5.2			2	ELA.5.2			3	ELA.5.1 4		
22	2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.2			2	ELA.5.7			2	ELA.5.7		
23	2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			3	ELA.5.3		
24	2	ELA.5.1			2	ELA.5.9			2	ELA.5.1			2	ELA.5.9			2	ELA.5.9			3	ELA.5.2		
25	2	ELA.5.9			2	ELA.5.9			2	ELA.5.2			3	ELA.5.9			2	ELA.5.9			2	ELA.5.9		
26	2	ELA.5.9			2	ELA.5.9			2	ELA.5.3			3	ELA.5.9			2	ELA.5.3			3	ELA.5.3		
27	2	ELA.5.1			3	ELA.5.2			2	ELA.5.1			3	ELA.5.2			2	ELA.5.8			3	ELA.5.2		
28	2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7			2	ELA.5.7		
29	2	ELA.5.1			2	ELA.5.9			2	ELA.5.8			3	ELA.5.9			2	ELA.5.2			3	ELA.5.9		
30	3	ELA.5.9			2	ELA.5.9			2	ELA.5.9			3	ELA.5.9			2	ELA.5.9			3	ELA.5.9		
31	2	ELA.5.7			2	ELA.5.1 0			2	ELA.5.1 0			1	ELA.5.3 9			2	ELA.5.1 0			2	ELA.5.7		
32	2	ELA.5.6			2	ELA.5.6			2	ELA.5.6			3	ELA.5.1 6			2	ELA.5.6			2	ELA.5.6		
33	2	ELA.5.4			2	ELA.5.4			2	ELA.5.6			3	ELA.5.1 6			3	ELA.5.1 6			3	ELA.5.1 6		
34	2	ELA.5.4			2	ELA.5.4			2	ELA.5.1 6			2	ELA.5.6			2	ELA.5.4			2	ELA.5.4		
35	2	ELA.5.4			2	ELA.5.4			2	ELA.5.4			2	ELA.5.4			2	ELA.5.4			2	ELA.5.4		
36	1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7		
37	2	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 6		
38	2	ELA.5.3 6			1	ELA.5.3 6			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 6			1	ELA.5.3 6		
39	1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7			1	ELA.5.3 7		
40	2	ELA.5.3 6	ELA.5.3 7		1	ELA.5.3 6	ELA.5.3 7		2	ELA.5.3 6	ELA.5.3 7		1	ELA.5.3 6	ELA.5.3 7		1	ELA.5.3 6	ELA.5.3 7		1	ELA.5.3 7	ELA.5.3 6	
41	3	ELA.5.2 0	ELA.5.3 8		3	ELA.5.2 0	ELA.5.3 8		4	ELA.5.2 0	ELA.5.2 3		3	ELA.5.2 1	ELA.5.3 8		3	ELA.5.2 0	ELA.5.3 8		3	ELA.5.3 8		
42	3	ELA.5.2 0	ELA.5.2 3	ELA.5.3 8	3	ELA.5.2 0	ELA.5.2 3		4	ELA.5.2 0	ELA.5.3 8		3	ELA.5.3 8	ELA.5.2 3	ELA.5.2 1	3	ELA.5.2 0	ELA.5.2 3		3	ELA.5.2 1	ELA.5.2 8	
Objective Pairwise Comparison: 0.63																								
Standard Pairwise Comparison: 0.97																								

Table 5.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV ELA 2019 Grade 5 B2

Low		Medium		High	
0		9		15	
LR.5					
LR.5.KID					
ELA.5.1	14(5)	15(4)	24(2)	27(2)	29(1)
ELA.5.2	29(1)	27(3)	25(1)	24(1)	21(3)
ELA.5.3	20(5)	23(1)	26(3)		
LR.5.CS					
ELA.5.7	28(6)	31(2)	23(5)	22(5)	15(1)
ELA.5.8	18(6)	19(6)	29(1)	27(1)	
ELA.5.9	26(3)	25(5)	29(3)	30(6)	20(1)
LR.5.IKI					
ELA.5.13					
ELA.5.14	21(3)				
IR.5					
IR.5.KID					
ELA.5.4	12(1)	8(1)	33(2)	34(4)	35(6)
ELA.5.5	6(6)	11(6)			
ELA.5.6	12(1)	10(1)	5(1)	32(5)	34(1)
IR.5.CS					
ELA.5.10	31(3)	13(5)	7(6)		
ELA.5.11	5(5)	8(3)	9(5)	12(1)	
ELA.5.12	12(1)	8(2)			
IR.5.IKI					

ELA.5.15								
ELA.5.16	10(5)	32(1)	33(3)	34(1)				
ELA.5.17	9(1)	12(2)						
WL.5								
WL.5.TTP								
ELA.5.20	35(1)	41(12)	42(12)					
ELA.5.21	41(3)	42(6)						
ELA.5.22								
WL.5.PDW								
ELA.5.23	41(3)	42(12)						
ELA.5.24								
ELA.5.25								
WL.5.RBPK								
ELA.5.26								
ELA.5.27								
ELA.5.28	42(3)							
WL.5.RW								
ELA.5.29								
WL.5.CSE								
ELA.5.36	37(6)	38(8)	1(5)	2(2)	3(6)	4(3)	40(12)	
ELA.5.37	4(3)	2(10)	1(1)	38(4)	39(6)	36(6)	40(12)	
WL.5.KL								
ELA.5.38	35(1)	41(15)	42(9)					
WL.5.VAU								
ELA.5.39	31(1)	13(1)						
ELA.5.40								
ELA.5.41								

Table 5.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 5 B2

Low		Medium		High	
3		9		15	
1 10199-3107	ELA.5.36:5	ELA.5.37:1			
2 10199-3108	ELA.5.36:2	ELA.5.37:10			
3 10199-3109	ELA.5.36:6				
4 10199-3110	ELA.5.36:3	ELA.5.37:3			
5 10199-8433	ELA.5.6:1	ELA.5.11:5			
6 10199-8434	ELA.5.5:6				
7 10199-8435	ELA.5.10:6				
8 10199-8436	ELA.5.4:1	ELA.5.11:3	ELA.5.12:2		
9 10199-8437	ELA.5.11:5	ELA.5.17:1			
10 10199-8439	ELA.5.6:1	ELA.5.16:5			
11 10199-9071	ELA.5.5:6				
12 10199-9072	ELA.5.4:1	ELA.5.6:1	ELA.5.11:1	ELA.5.12:1	ELA.5.17:2
13 10199-9074	ELA.5.10:5	ELA.5.39:1			
14 10199-11018	ELA.5.1:5	ELA.5.9:1			
15 10199-11020	ELA.5.1:4	ELA.5.2:1	ELA.5.7:1		
16 10199-11021	ELA.5.2:6				
17 10199-11024	ELA.5.7:6				
18 10199-11025	ELA.5.8:6				
19 10199-11026	ELA.5.8:6				
20 10199-11027	ELA.5.3:5	ELA.5.9:1			
21 10199-11028	ELA.5.2:3	ELA.5.14:3			
22 10199-11030	ELA.5.2:1	ELA.5.7:5			
23 10199-11031	ELA.5.3:1	ELA.5.7:5			
24 10199-4678	ELA.5.1:2	ELA.5.2:1	ELA.5.9:3		
25 10199-4679	ELA.5.2:1	ELA.5.9:5			
26 10199-4680	ELA.5.3:3	ELA.5.9:3			
27 10199-4681	ELA.5.1:2	ELA.5.2:3	ELA.5.8:1		
28 10199-4682	ELA.5.7:6				
29 10199-4683	ELA.5.1:1	ELA.5.2:1	ELA.5.8:1	ELA.5.9:3	
30 10199-4684	ELA.5.9:6				

31 10199-7078	ELA.5.7:2	ELA.5.10:3	ELA.5.39:1		
32 10199-7084	ELA.5.6:5	ELA.5.16:1			
33 10199-7086	ELA.5.4:2	ELA.5.6:1	ELA.5.16:3		
34 10199-7092	ELA.5.4:4	ELA.5.6:1	ELA.5.16:1		
35 10199-7097	ELA.5.4:6	ELA.5.20:1	ELA.5.38:1		
36 10199-1366	ELA.5.37:6				
37 10199-1367	ELA.5.36:6				
38 10199-1370	ELA.5.36:8	ELA.5.37:4			
39 10199-1371	ELA.5.37:6				
40 CONVENTIONS10199- 14172	ELA.5.36:12	ELA.5.37:12			
41 ELABORATION10199- 14172	ELA.5.20:12	ELA.5.21:3	ELA.5.23:3	ELA.5.38:15	
42 ORGANIZATION10199 -14172	ELA.5.20:12	ELA.5.21:6	ELA.5.23:12	ELA.5.28:3	ELA.5.38:9

Table 5.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 5 B2

Low DOK		Matched DOK		High DOK

LR.5								
LR.5.KID								
ELA.5.1: [2]	14:(5)[2]	15:(4)[2]	24:(2)[2]	27:(2)[2]	29:(1)[2]			
ELA.5.2: [3]	15:(1)[2]	16:(6)[3]	21:(3)[3]	22:(1)[2]	24:(1)[3]	25:(1)[2]	27:(3)[3]	29:(1)[2]
ELA.5.3: [3]	20:(5)[2]	23:(1)[3]	26:(3)[2]					
LR.5.CS								
ELA.5.7: [2]	15:(1)[2]	17:(6)[2]	22:(5)[2]	23:(5)[2]	28:(6)[2]	31:(2)[2]		
ELA.5.8: [2]	18:(6)[2]	19:(6)[2]	27:(1)[2]	29:(1)[2]				
ELA.5.9: [2]	14:(1)[3]	20:(1)[2]	24:(3)[2]	25:(5)[2]	26:(3)[2]	29:(3)[3]	30:(6)[2]	
LR.5.IKI								
ELA.5.13								
ELA.5.14: [3]	21:(3)[2]							
IR.5								
IR.5.KID								
ELA.5.4: [2]	8:(1)[2]	12:(1)[3]	33:(2)[2]	34:(4)[2]	35:(6)[2]			
ELA.5.5: [2]	6:(6)[2]	11:(6)[2]						
ELA.5.6: [2]	5:(1)[2]	10:(1)[2]	12:(1)[2]	32:(5)[2]	33:(1)[2]	34:(1)[2]		
IR.5.CS								
ELA.5.10: [2]	7:(6)[2]	13:(5)[2]	31:(3)[2]					
ELA.5.11: [2]	5:(5)[3]	8:(3)[2]	9:(5)[2]	12:(1)[2]				
ELA.5.12: [3]	8:(2)[2]	12:(1)[3]						
IR.5.IKI								
ELA.5.15								
ELA.5.16: [3]	10:(5)[3]	32:(1)[3]	33:(3)[3]	34:(1)[2]				
ELA.5.17: [3]	9:(1)[3]	12:(2)[2]						
WL.5								
WL.5.TTP								
ELA.5.20: [3]	35:(1)[2]	41:(12)[3]	42:(12)[3]					

ELA.5.21: [3]	41:(3)[3]	42:(6)[3]						
ELA.5.22								
WL.5.PDW								
ELA.5.23: [3]	41:(3)[4]	42:(12)[3]						
ELA.5.24								
ELA.5.25								
WL.5.RBPK								
ELA.5.26								
ELA.5.27								
ELA.5.28: [3]	42:(3)[3]							
WL.5.RW								
ELA.5.29								
WL.5.CSE								
ELA.5.36: [2]	1:(5)[1]	2:(2)[1]	3:(6)[1]	4:(3)[1]	37:(6)[1]	38:(8)[1]	40:(12)[1]	
ELA.5.37: [1]	1:(1)[1]	2:(10)[1]	4:(3)[1]	36:(6)[1]	38:(4)[1]	39:(6)[1]	40:(12)[1]	
WL.5.KL								
ELA.5.38: [2]	35:(1)[2]	41:(15)[3]	42:(9)[3]					
WL.5.VAU								
ELA.5.39: [2]	13:(1)[2]	31:(1)[1]						
ELA.5.40								
ELA.5.41								

Table 5.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 CONF WV Grade 5B2 ELA

Low	Medium	High
2.4	7.2	12

1 10199-3107	Exact:5	
2 10199-3108	Exact:10	
3 10199-3109	Exact:5	
4 10199-3110	Exact:5	
5 10199-8433	Exact:3	Partial:2
6 10199-8434	Exact:5	
7 10199-8435	Exact:4	Minimal:1
8 10199-8436	Exact:4	Partial:1
9 10199-8437	Exact:4	Partial:1
10 10199-8439	Exact:5	
11 10199-9071	Exact:5	
12 10199-9072	Exact:5	
13 10199-9074	Exact:5	
14 10199-11018	Exact:5	
15 10199-11020	Exact:5	
16 10199-11021	Exact:5	
17 10199-11024	Exact:5	
18 10199-11025	Exact:5	
19 10199-11026	Exact:5	
20 10199-11027	Exact:5	
21 10199-11028	Exact:5	
22 10199-11030	Exact:5	
23 10199-11031	Exact:5	
24 10199-4678	Exact:5	
25 10199-4679	Exact:4	Partial:1
26 10199-4680	Exact:5	
27 10199-4681	Exact:4	Partial:1
28 10199-4682	Exact:5	
29 10199-4683	Exact:4	Partial:1
30 10199-4684	Exact:4	Partial:1

31 10199-7078	Exact:4	Partial:1
32 10199-7084	Exact:5	
33 10199-7086	Exact:3	Partial:2
34 10199-7092	Exact:5	
35 10199-7097	Exact:5	
36 10199-1366	Exact:5	
37 10199-1367	Exact:5	
38 10199-1370	Exact:10	
39 10199-1371	Exact:5	
40 CONVENTIONS10199-14172	Partial:2	Minimal:8
41 ELABORATION10199-14172	Partial:12	Minimal:3
42 ORGANIZATION10199-14172	Partial:12	Minimal:3

Grade 6 Batch 1 ELA West Virginia

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 6 B1 Number of Assessment Items - 41*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.6 Grade 6 Literary Reading	3	9	2 3	6 2	75 25	14.2	1.64	YES
IR.6 Grade 6 Informational Reading	3	9	2 3	6 3	66.67 33.33	19.8	0.84	YES
WL.6 Grade 6 Writing and Language	7	16	1 2 3 4	2 5 8 1	12.5 31.25 50 6.25	26.8	1.79	YES
Total	13	34	1 2 3 4	2 17 13 1	6 52 39 3	60.8	2.17	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 6 B1 Number of Assessment Items - 41*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.6 Grade 6 Literary Reading	3	9	14.2	1.64	0	0	98.57	3	1.43	3	YES
IR.6 Grade 6 Informational Reading	3	9	19.8	0.84	20.48	12	79.52	12	0	0	YES
WL.6 Grade 6 Writing and Language	7	16	26.8	1.79	0	0	74.36	18	25.64	18	YES
Total	13	34	60.8	2.17	6.58	3.9	81.91	7.5	11.51	6.8	

*Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Five Reviewers
WV ELA 2019 Grade 6 B1 Number of Assessment Items - 41*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.6 Grade 6 Literary Reading	3	9	14.2	1.64	6.4	0.55	71.11	6.09	YES	31	2	0.75	0.04	YES
IR.6 Grade 6 Informational Reading	3	9	19.8	0.84	7.8	0.45	86.67	4.97	YES	43	1	0.77	0.03	YES
WL.6 Grade 6 Writing and Language	7	16	26.8	1.79	5.2	0.45	32.5	2.8	NO	27	2	0.66	0.03	WEAK
Total	13	34	60.8	2.17	6.5	1.3	63.43	28		34	8	0.73	0.06	

*Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Five Reviewers WV ELA 2019 Grade 6 B1
Number of Assessment Items - 41*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.6 Grade 6 Literary Reading	YES	YES	YES	YES
IR.6 Grade 6 Informational Reading	YES	YES	YES	YES
WL.6 Grade 6 Writing and Language	YES	YES	NO	WEAK

Table 6.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 6 B1

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1	2	2	2	2	2
2	2	2	2	2	2
3	2	2	2	2	2
4	2	2	2	2	2
5	2	2	2	2	2
6	2	2	2	2	2
7	2	2	2	2	2
8	2	2	2	2	2
9	2	2	2	2	2
10	2	2	2	2	2
11	2	2	2	2	2
12	2	2	2	2	2
13	2	2	2	2	2
14	2	2	2	2	2
15	2	2	2	2	2
16	2	2	2	2	3
17	2	2	2	2	2
18	1	1	2	1	1
19	2	1	2	1	1
20	2	1	2	1	1
21	1	1	2	2	2
22	1	1	2	2	2
23	2	2	2	2	2
24	2	2	2	2	2
25	2	2	2	2	3
26	2	3	2	2	3
27	2	2	2	2	3
28	2	2	2	2	2
29	2	2	2	2	2
30	2	2	2	2	2
31	2	1	2	1	1

32	2	1	2	1	1
33	1	2	2	1	1
34	2	2	2	2	2
35	2	2	2	2	2
36	2	2	2	2	2
37	2	2	2	2	2
38	2	2	2	2	2
39	1	1	1	1	1
40	3	3	3	3	3
41	3	3	3	3	3

Intraclass correlation - .9174

Pairwise Comparison - 0.84

Table 6.6
DOK Levels and Objectives Code by Each Reviewer
WV ELA 2019 Grade 6 B1

Number of Reviewers: Five

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10		
2	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11		
3	2	ELA.6.4			2	ELA.6.16			2	ELA.6.4			2	ELA.6.16			2	ELA.6.4		
4	2	ELA.6.5			2	ELA.6.5			2	ELA.6.5			2	ELA.6.2			2	ELA.6.5		
5	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.11			2	ELA.6.4		
6	2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.4	ELA.6.10	
7	2	ELA.6.12			2	ELA.6.12			2	ELA.6.12	ELA.6.5		2	ELA.6.12			2	ELA.6.4		
8	2	ELA.6.16			2	ELA.6.16			2	ELA.6.4			2	ELA.6.16			2	ELA.6.16		
9	2	ELA.6.7			2	ELA.6.2			2	ELA.6.7			2	ELA.6.7			2	ELA.6.7		
10	2	ELA.6.9			2	ELA.6.3			2	ELA.6.9			2	ELA.6.7			2	ELA.6.3		
11	2	ELA.6.7			2	ELA.6.7			2	ELA.6.7			2	ELA.6.7			2	ELA.6.7		
12	2	ELA.6.1			2	ELA.6.1			2	ELA.6.2			2	ELA.6.1			2	ELA.6.1		
13	2	ELA.6.7			2	ELA.6.40			2	ELA.6.7			2	ELA.6.7			2	ELA.6.7		
14	2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.8			2	ELA.6.1		
15	2	ELA.6.1			2	ELA.6.9			2	ELA.6.1	ELA.6.9		2	ELA.6.3			2	ELA.6.1		
16	2	LR.6			2	LR.6			2	LR.6			2	LR.6			3	LR.6		
17	2	ELA.6.7			2	ELA.6.7			2	ELA.6.7			2	ELA.6.7			2	ELA.6.7		
18	1	ELA.6.37			1	ELA.6.37			2	ELA.6.37			1	ELA.6.36			1	ELA.6.37		
19	2	ELA.6.36			1	ELA.6.36			2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
20	2	ELA.6.36			1	ELA.6.36			2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
21	1	ELA.6.4			1	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4		
22	1	ELA.6.4			1	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4		
23	2	ELA.6.11	ELA.6.5		2	ELA.6.5	ELA.6.11		2	ELA.6.5	ELA.6.11		2	ELA.6.11	ELA.6.5		2	ELA.6.5	ELA.6.11	
24	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11		
25	2	ELA.6.15			2	ELA.6.15			2	ELA.6.15			2	ELA.6.15			3	ELA.6.15		
26	2	ELA.6.16			3	ELA.6.16			2	ELA.6.12			2	ELA.6.16			3	ELA.6.16		
27	2	ELA.6.17			2	ELA.6.17			2	ELA.6.17			2	ELA.6.17			3	ELA.6.17		
28	2	ELA.6.5			2	ELA.6.5			2	ELA.6.5			2	ELA.6.5			2	ELA.6.5		

29	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11	ELA.6.12		2	ELA.6.12			2	ELA.6.11		
30	2	ELA.6.12			2	ELA.6.12			2	ELA.6.12			2	ELA.6.12			2	ELA.6.12		
31	2	ELA.6.36			1	ELA.6.36			2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
32	2	ELA.6.36			1	ELA.6.36			2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
33	1	ELA.6.37			2	ELA.6.37			2	ELA.6.37			1	ELA.6.37			1	ELA.6.37		
34	2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.8			2	ELA.6.1		
35	2	ELA.6.8			2	ELA.6.8			2	ELA.6.8			2	ELA.6.8			2	ELA.6.9		
36	2	ELA.6.4			2	ELA.6.9			2	ELA.6.9	ELA.6.1		2	ELA.6.12			2	ELA.6.2		
37	2	ELA.6.2			2	ELA.6.2			2	ELA.6.1	ELA.6.2		2	ELA.6.2			2	ELA.6.8		
38	2	ELA.6.2			2	ELA.6.2			2	ELA.6.2			2	ELA.6.2			2	ELA.6.2		
39	1	ELA.6.36	ELA.6.37		1	ELA.6.36	ELA.6.37		1	ELA.6.36	ELA.6.37		1	ELA.6.36	ELA.6.37		1	ELA.6.36	ELA.6.37	
40	3	ELA.6.21			3	ELA.6.21	ELA.6.28		3	ELA.6.21			3	ELA.6.21	ELA.6.23	ELA.6.38	3	ELA.6.21		
41	3	ELA.6.21	ELA.6.23	ELA.6.38	3	ELA.6.21	ELA.6.28	ELA.6.38	3	ELA.6.21	ELA.6.23	ELA.6.38	3	ELA.6.21			3	ELA.6.21	ELA.6.38	ELA.6.23
Objective Pairwise Comparison: 0.72																				
Standard Pairwise Comparison: 0.97																				

Table 6.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers) WV ELA 2019 Grade 6 B1

	Low		Medium			High		
	0		9				15	
LR.6	16(5)							
LR.6.KID								
ELA.6.1	14(4)	12(4)	34(4)	15(3)	37(1)	36(1)		
ELA.6.2	38(5)	36(1)	12(1)	9(1)	4(1)	37(4)		
ELA.6.3	10(2)	15(1)						
LR.6.CS								
ELA.6.7	10(1)	9(4)	13(4)	11(5)	17(5)			
ELA.6.8	14(1)	34(1)	35(4)	37(1)				
ELA.6.9	35(1)	36(2)	10(2)	15(2)				
LR.6.IKI								
ELA.6.13								
ELA.6.14								
IR.6								
IR.6.KID								
ELA.6.4	3(3)	5(4)	6(1)	7(1)	8(1)	36(1)	21(5)	22(5)
ELA.6.5	28(5)	4(4)	23(5)	7(1)				
ELA.6.6								
IR.6.CS								
ELA.6.10	1(5)	6(5)						
ELA.6.11	5(1)	2(5)	29(4)	24(5)	23(5)			
ELA.6.12	26(1)	30(5)	36(1)	7(4)	29(2)			
IR.6.IKI								

ELA.6.15	25(5)						
ELA.6.16	26(4)	8(4)	3(2)				
ELA.6.17	27(5)						
WL.6							
WL.6.TTP							
ELA.6.20							
ELA.6.21	40(15)	41(15)					
ELA.6.22							
WL.6.PDW							
ELA.6.23	40(3)	41(9)					
ELA.6.24							
ELA.6.25							
WL.6.RBPK							
ELA.6.26							
ELA.6.27							
ELA.6.28	40(3)	41(3)					
WL.6.RW							
ELA.6.29							
WL.6.CSE							
ELA.6.36	39(10)	19(10)	20(10)	31(5)	32(10)	18(1)	
ELA.6.37	18(4)	33(10)	39(10)				
WL.6.KL							
ELA.6.38	40(3)	41(12)					
WL.6.VAU							
ELA.6.39							
ELA.6.40	13(1)						
ELA.6.41							

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

WV ELA 2019 Grade 6 B1

Low		Medium		High
3		9		15
1 10199-3422	ELA.6.10:5			
2 10199-3429	ELA.6.11:5			
3 10199-3590	ELA.6.4:3	ELA.6.16:2		
4 10199-3599	ELA.6.2:1	ELA.6.5:4		
5 10199-3608	ELA.6.4:4	ELA.6.11:1		
6 10199-4843	ELA.6.4:1	ELA.6.10:5		
7 10199-4848	ELA.6.4:1	ELA.6.5:1	ELA.6.12:4	
8 10199-6448	ELA.6.4:1	ELA.6.16:4		
9 10199-7184	ELA.6.2:1	ELA.6.7:4		
10 10199-7185	ELA.6.3:2	ELA.6.7:1	ELA.6.9:2	
11 10199-7186	ELA.6.7:5			
12 10199-7187	ELA.6.1:4	ELA.6.2:1		
13 10199-7190	ELA.6.7:4	ELA.6.40:1		
14 10199-7191	ELA.6.1:4	ELA.6.8:1		
15 10199-7192	ELA.6.1:3	ELA.6.3:1	ELA.6.9:2	
16 10199-7194	LR.6:5			
17 10199-7200	ELA.6.7:5			
18 10199-1396	ELA.6.36:1	ELA.6.37:4		
19 10199-1397	ELA.6.36:10			
20 10199-1399	ELA.6.36:10			
21 10199-4117	ELA.6.4:5			
22 10199-4118	ELA.6.4:5			
23 10199-4119	ELA.6.5:5	ELA.6.11:5		
24 10199-4120	ELA.6.11:5			
25 10199-4122	ELA.6.15:5			
26 10199-4124	ELA.6.12:1	ELA.6.16:4		

27 10199-4183	ELA.6.17:5				
28 10199-4847	ELA.6.5:5				
29 10199-4182	ELA.6.11:4	ELA.6.12:2			
30 10199-4856	ELA.6.12:5				
31 REP10199-1356	ELA.6.36:5				
32 REP10199-1359	ELA.6.36:10				
33 REP10199-1363	ELA.6.37:10				
34 10199-13958	ELA.6.1:4	ELA.6.8:1			
35 REP10199-13964	ELA.6.8:4	ELA.6.9:1			
36 10199-13965	ELA.6.1:1	ELA.6.2:1	ELA.6.9:2	ELA.6.4:1	ELA.6.12:1
37 REP10199-13970	ELA.6.1:1	ELA.6.2:4	ELA.6.8:1		
38 REP10199-13982	ELA.6.2:5				
39 CONVENTIONS10199-14166	ELA.6.36:10	ELA.6.37:10			
40 ELABORATION10199-14166	ELA.6.21:15	ELA.6.23:3	ELA.6.28:3	ELA.6.38:3	
41 ORGANIZATION10199-14166	ELA.6.21:15	ELA.6.23:9	ELA.6.28:3	ELA.6.38:12	

Table 6.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 6 B1

Low DOK		Matched DOK		High DOK

LR.6: [3]	16:(5)[2]							
LR.6.KID								
ELA.6.1: [2]	12:(4)[2]	14:(4)[2]	15:(3)[2]	34:(4)[2]	36:(1)[2]	37:(1)[2]		
ELA.6.2: [2]	4:(1)[2]	9:(1)[2]	12:(1)[2]	36:(1)[2]	37:(4)[2]	38:(5)[2]		
ELA.6.3: [2]	10:(2)[2]	15:(1)[2]						
LR.6.CS								
ELA.6.7: [2]	9:(4)[2]	10:(1)[2]	11:(5)[2]	13:(4)[2]	17:(5)[2]			
ELA.6.8: [2]	14:(1)[2]	34:(1)[2]	35:(4)[2]	37:(1)[2]				
ELA.6.9: [2]	10:(2)[2]	15:(2)[2]	35:(1)[2]	36:(2)[2]				
LR.6.IKI								
ELA.6.13								
ELA.6.14								
IR.6								
IR.6.KID								
ELA.6.4: [2]	3:(3)[2]	5:(4)[2]	6:(1)[2]	7:(1)[2]	8:(1)[2]	21:(5)[2]	22:(5)[2]	36:(1)[2]
ELA.6.5: [2]	4:(4)[2]	7:(1)[2]	23:(5)[2]	28:(5)[2]				
ELA.6.6								
IR.6.CS								
ELA.6.10: [2]	1:(5)[2]	6:(5)[2]						
ELA.6.11: [2]	2:(5)[2]	5:(1)[2]	23:(5)[2]	24:(5)[2]	29:(4)[2]			
ELA.6.12: [2]	7:(4)[2]	26:(1)[2]	29:(2)[2]	30:(5)[2]	36:(1)[2]			
IR.6.IKI								
ELA.6.15: [3]	25:(5)[2]							
ELA.6.16: [3]	3:(2)[2]	8:(4)[2]	26:(4)[2]					
ELA.6.17: [3]	27:(5)[2]							
WL.6								
WL.6.TTP								
ELA.6.20								

ELA.6.21: [3]	40:(15)[3]	41:(15)[3]					
ELA.6.22							
WL.6.PDW							
ELA.6.23: [3]	40:(3)[3]	41:(9)[3]					
ELA.6.24							
ELA.6.25							
WL.6.RBPK							
ELA.6.26							
ELA.6.27							
ELA.6.28: [3]	40:(3)[3]	41:(3)[3]					
WL.6.RW							
ELA.6.29							
WL.6.CSE							
ELA.6.36: [1]	18:(1)[1]	19:(10)[1]	20:(10)[1]	31:(5)[1]	32:(10)[1]	39:(10)[1]	
ELA.6.37: [1]	18:(4)[1]	33:(10)[1]	39:(10)[1]				
WL.6.KL							
ELA.6.38: [2]	40:(3)[3]	41:(12)[3]					
WL.6.VAU							
ELA.6.39							
ELA.6.40: [2]	13:(1)[2]						
ELA.6.41							

Table 6.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
CONF WV Grade 6B1 ELA

Low		Medium		High	
3.6		10.8		18	
1	10199-3422	Exact:3	Partial:3		
2	10199-3429	Exact:5	Partial:1		
3	10199-3590	Partial:5	Negligible:1		
4	10199-3599	Exact:6			
5	10199-3608	Exact:5	Partial:1		
6	10199-4843	Exact:1	Partial:3	Minimal:2	
7	10199-4848	Exact:5	Partial:1		
8	10199-6448	Exact:5	Partial:1		
9	10199-7184	Exact:2	Partial:4		
10	10199-7185	Exact:2	Partial:3		
11	10199-7186	Exact:3	Partial:3		
12	10199-7187	Partial:5	Negligible:1		
13	10199-7190	Exact:2	Partial:3	Minimal:1	
14	10199-7191	Exact:5	Minimal:1		
15	10199-7192	Exact:5	Minimal:1		
16	10199-7194	Exact:2	Partial:1	Minimal:1	Negligible:2
17	10199-7200	Exact:5	Partial:1		
18	10199-1396	Exact:6			
19	10199-1397	Exact:12			
20	10199-1399	Exact:12			
21	10199-4117	Exact:6			
22	10199-4118	Exact:6			
23	10199-4119	Exact:4	Partial:1	Minimal:1	
24	10199-4120	Exact:3	Partial:1	Minimal:2	
25	10199-4122	Exact:4	Partial:2		
26	10199-4124	Exact:4	Partial:2		
27	10199-4183	Exact:4	Partial:2		
28	10199-4847	Exact:5			
29	10199-4182	Exact:3	Partial:1	Minimal:1	
30	10199-4856	Exact:6			
31	REP10199-1356	Exact:6			

32 REP10199-1359	Exact:12			
33 REP10199-1363	Exact:12			
34 10199-13958	Exact:5	Partial:1		
35 REP10199-13964	Exact:6			
36 10199-13965	Exact:6			
37 REP10199-13970	Exact:6			
38 REP10199-13982	Exact:2	Partial:2	Minimal:1	Negligible:1
39 CONVENTIONS10199-14166	Exact:2	Partial:2	Minimal:6	Negligible:2
40 ELABORATION10199-14166	Exact:18			
41 ORGANIZATION10199-14166	Exact:3	Partial:12	Negligible:3	

Grade 6 Batch 2 ELA West Virginia - Elementary Panel

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 6 B2 ELEM Number of Assessment Items - 41*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.6 Grade 6 Literary Reading	3	8	2 3	6 2	75 25	12.57	6.05	YES
IR.6 Grade 6 Informational Reading	3	9	2 3	6 3	66.67 33.33	17	5.72	YES
WL.6 Grade 6 Writing and Language	7	16	1 2 3 4	2 5 8 1	12.5 31.25 50 6.25	25.43	11.27	YES
Total	13	33	1 2 3 4	2 17 13 1	6 52 39 3	55	22.02	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Six Reviewers WV ELA 2019 Grade 6 B2 ELEM Number of Assessment Items - 41*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.6 Grade 6 Literary Reading	3	8	12.57	6.05	10.01	8	78.71	14	11.28	12	YES
IR.6 Grade 6 Informational Reading	3	9	17	5.72	26.99	11	70.47	15	2.54	5	YES
WL.6 Grade 6 Writing and Language	7	16	25.43	11.27	22.99	37	57.65	30	19.36	18	YES
Total	13	33	55	22.02	15.58	10.7	71.17	9.2	13.25	8	

*Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers
WV ELA 2019 Grade 6 B2 ELEM Number of Assessment Items - 41*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.6 Grade 6 Literary Reading	3	8	12.57	6.05	6	2.65	75	33.07	YES	26	13	0.88	0.06	YES
IR.6 Grade 6 Informational Reading	3	9	17	5.72	7	1.29	77.78	14.34	YES	46	17	0.8	0.10	YES
WL.6 Grade 6 Writing and Language	7	16	25.43	11.27	5.43	2.07	33.93	12.94	NO	28	5	0.77	0.13	YES
Total	13	33	55	22.02	6.1	0.8	62.24	25		33	11	0.82	0.06	

*Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV ELA 2019 Grade 6 B2 ELEM
Number of Assessment Items - 41*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.6 Grade 6 Literary Reading	YES	YES	YES	YES
IR.6 Grade 6 Informational Reading	YES	YES	YES	YES
WL.6 Grade 6 Writing and Language	YES	YES	NO	YES

Table 6.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 6 B2 ELEM

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6	Reviewer 7
1	2	2	2	2	2	2	2
2	1	2	2	1	1	2	2
3	2	2	2	2	2	2	2
4	2	2	2	2	1	2	2
5	2	2	2	2	2	2	2
6	2	1	2	2	2	2	3
7	2	2	2	2		2	2
8	2	1	1	2		2	1
9	2	3	2	2		2	3
10	3	3	2	2		3	2
11	2	2	2	2		2	1
12	2	2	2	2		2	2
13	2	2	2	2		2	2
14	2	3	2	2		2	2
15	2	2	2	2		2	2
16	2	1	1	1		2	2
17	2	2	3	2		3	3
18	2	1	1	1		1	1
19	2	1	1	1		1	1
20	2	1	1	1		1	1
21	2	1	2	2		2	2
22	2	2	2	2		2	1
23	2	2	2	2		2	1
24	2	1	2	2		2	1
25	2	2	3	2		2	2
26	2	3	2	2		2	2
27	2	1	2	2		2	2
28	2	3	2	2		2	2
29	2	2	2	2		2	2
30	2	2	2	2		2	2

31	2	1	1	1		1	1
32	2	1	1	1		1	1
33	2	1	1	1		1	1
34	2	2	2	2		2	2
35	2	3	2	2		3	3
36	2	2	2	2		2	2
37	1	2	2	2		2	2
38	1	2	2	2		1	2
39	3	1	1	1		1	1
40	3	3	3	3		3	3
41	3	3	3	3		3	3

Intraclass correlation - .8436

Pairwise Comparison - 0.72

Table 6.6 DOK Levels and Objectives Code by Each Reviewer WV ELA 2019 Grade 6 B2 ELEM

Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.6.5	ELA.6.1		2	ELA.6.2			2	ELA.6.1			2	ELA.6.5	ELA.6.1		2	ELA.6.5			2	ELA.6.5			2	ELA.6.1		
2	1	ELA.6.1	0		2	ELA.6.1	0		2	ELA.6.1	0		1	ELA.6.3	9		1	ELA.6.3	9		2	ELA.6.1	0		2	ELA.6.1	0	
3	2	ELA.6.1	0		2	ELA.6.1	0		2	ELA.6.1	0		2	ELA.6.1	0		2	ELA.6.1	0		2	ELA.6.1	0		2	ELA.6.1	0	
4	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			1	ELA.6.4			2	ELA.6.4			2	ELA.6.4		
5	2	ELA.6.1	2		2	ELA.6.1	2		2	ELA.6.1	2		2	ELA.6.1	2		2	ELA.6.1	2		2	ELA.6.1	2		2	ELA.6.1	2	
6	2	ELA.6.1	1		1	ELA.6.6			2	ELA.6.1	1		2	ELA.6.1	1		2	ELA.6.1	1		2	ELA.6.1	1		3	ELA.6.1	2	
7	2	ELA.6.4			2	ELA.6.4			2	ELA.6.1	1		2	ELA.6.4							2	ELA.6.4			2	ELA.6.2		
8	2	ELA.6.4			1	ELA.6.4			1	ELA.6.4			2	ELA.6.5							2	ELA.6.4			1	ELA.6.4		
9	2	ELA.6.9			3	ELA.6.9			2	ELA.6.9			2	ELA.6.8							2	ELA.6.7			3	ELA.6.9		
10	3	ELA.6.2			3	ELA.6.2			2	ELA.6.2	ELA.6.8		2	ELA.6.2							3	ELA.6.2	ELA.6.1		2	ELA.6.2	ELA.6.1	
11	2	ELA.6.7			2	ELA.6.7			2	ELA.6.7			2	ELA.6.7										1	ELA.6.7			
12	2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.1							2	ELA.6.1			2	ELA.6.1		
13	2	ELA.6.8			2	ELA.6.8			2	ELA.6.8			2	ELA.6.8							2	ELA.6.8			2	ELA.6.3		
14	2	ELA.6.1			3	ELA.6.7			2	ELA.6.8			2	ELA.6.1							2	ELA.6.1			2	ELA.6.1		
15	2	ELA.6.9			2	ELA.6.9			2	ELA.6.9			2	ELA.6.9							2	ELA.6.9			2	ELA.6.9		
16	2	ELA.6.4	0		1	ELA.6.4	0		1	ELA.6.4	0		1	ELA.6.4	0						2	ELA.6.4	0		2	ELA.6.4	0	
17	2	ELA.6.1	4		2	ELA.6.1	4		3	ELA.6.1	4		2	ELA.6.1	4						3	ELA.6.1	4		3	ELA.6.1	4	
18	2	ELA.6.3	6		1	ELA.6.3	6		1	ELA.6.3	6		1	ELA.6.3	6						1	ELA.6.3	6		1	ELA.6.4	0	
19	2	ELA.6.3	6		1	ELA.6.3	6		1	ELA.6.3	6		1	ELA.6.3	6						1	ELA.6.3	6		1	ELA.6.4	1	
20	2	ELA.6.3	7		1	ELA.6.3	7		1	ELA.6.3	7		1	ELA.6.3	7						1	ELA.6.3	6	ELA.6.3	7	1	ELA.6.4	1
21	2	ELA.6.4			1	ELA.6.4			2	ELA.6.1	1		2	ELA.6.4							2	ELA.6.4			2	ELA.6.1		
22	2	ELA.6.4			2	ELA.6.5			2	ELA.6.5			2	ELA.6.4							2	ELA.6.1	1	ELA.6.4	1	ELA.6.4		
23	2	ELA.6.4			2	ELA.6.4			2	ELA.6.6			2	ELA.6.4							2	ELA.6.1	2	ELA.6.4		1	ELA.6.8	

24	2	ELA.6.1 0		1	ELA.6.1 0		2	ELA.6.1 0		2	ELA.6.1 0			2	ELA.6.1 0		1	ELA.6.7		
25	2	ELA.6.4		2	ELA.6.1 6	ELA.6.4	3	ELA.6.1 6		2	ELA.6.4	ELA.6.1 6		2	ELA.6.4	ELA.6.1 6	2	ELA.6.4	ELA.6.1 6	
26	2	ELA.6.1 1	ELA.6.1 6	3	ELA.6.1 1		2	ELA.6.1 1	ELA.6.1 6	2	ELA.6.1 1	ELA.6.1 6		2	ELA.6.1 6	ELA.6.1 1	2	ELA.6.1 1	ELA.6.1 6	
27	2	ELA.6.1 2		1	ELA.6.4		2	ELA.6.1 7		2	ELA.6.4			2	ELA.6.4		2	ELA.6.4		
28	2	ELA.6.1 2		3	ELA.6.1 2		2	ELA.6.1 2		2	ELA.6.1 2			2	ELA.6.1 2		2	ELA.6.9		
29	2	ELA.6.1 5		2	ELA.6.1 5		2	ELA.6.1 5		2	ELA.6.1 7			2	ELA.6.1 5		2	ELA.6.1 5		
30	2	ELA.6.1 6		2	ELA.6.1 6		2	ELA.6.1 6		2	ELA.6.1 6			2	ELA.6.1 6		2	ELA.6.1 6		
31	2	ELA.6.3 6		1	ELA.6.3 6		1	ELA.6.3 6		1	ELA.6.3 6			1	ELA.6.3 6		1	ELA.6.4 1		
32	2	ELA.6.3 7		1	ELA.6.3 7		1	ELA.6.3 7		1	ELA.6.3 7			1	ELA.6.3 6	ELA.6.3 7	1	ELA.6.4 1		
33	2	ELA.6.3 7		1	ELA.6.3 7		1	ELA.6.3 7		1	ELA.6.3 7			1	ELA.6.3 7		1	ELA.6.4 1		
34	2	ELA.6.8		2	ELA.6.8		2	ELA.6.8		2	ELA.6.8			2	ELA.6.8		2	ELA.6.8		
35	2	ELA.6.7		3	ELA.6.7		2	ELA.6.7		2	ELA.6.7			3	ELA.6.7		3	ELA.6.7		
36	2	ELA.6.2		2	ELA.6.8		2	ELA.6.2		2	ELA.6.2			2	ELA.6.3		2	ELA.6.2	ELA.6.8	
37	1	ELA.6.2		2	ELA.6.3		2	ELA.6.3		2	ELA.6.3			2	ELA.6.3		2	ELA.6.3		
38	1	ELA.6.3		2	ELA.6.3		2	ELA.6.3		2	ELA.6.3			1	ELA.6.3		2	ELA.6.3		
39	3	ELA.6.3 6	ELA.6.3 7	1	ELA.6.3 6	ELA.6.3 7	1	ELA.6.3 6	ELA.6.3 7	ELA.6.3 8	1	ELA.6.3 6	ELA.6.3 7	1	ELA.6.3 6	ELA.6.3 7	1	ELA.6.3 8	ELA.6.3 6	ELA.6.3 7
40	3	ELA.6.2 1	ELA.6.2 8	3	ELA.6.2 1		3	ELA.6.3 8	ELA.6.2 1	3	ELA.6.2 1	ELA.6.3 8		3	ELA.6.2 1	ELA.6.3 8	3	ELA.6.2 1		
41	3	ELA.6.2 1		3	ELA.6.2 1	ELA.6.2 3	ELA.6.3 8	3	ELA.6.3 8	ELA.6.2 1	ELA.6.2 3	3	ELA.6.2 1	ELA.6.3 8	ELA.6.2 3	3	ELA.6.2 1	ELA.6.3 8	ELA.6.2 3	ELA.6.3 8

Objective Pairwise Comparison: 0.65

Standard Pairwise Comparison: 0.94

Table 6.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV ELA 2019 Grade 6 B2 ELEM

	Low		Medium		High			
	0		10.8				18	
LR.6								
LR.6.KID								
ELA.6.1	12(6)	14(4)	21(1)	10(2)				
ELA.6.2	10(6)	1(1)	7(1)	37(1)	36(4)			
ELA.6.3	36(1)	37(5)	38(6)	13(1)				
LR.6.CS								
ELA.6.7	11(6)	14(1)	9(1)	24(1)	35(6)			
ELA.6.8	34(6)	9(1)	23(1)	14(1)	13(5)	10(1)	36(2)	
ELA.6.9	15(6)	9(4)	28(1)					
LR.6.IKI								
ELA.6.13								
ELA.6.14	17(6)							
IR.6								
IR.6.KID								
ELA.6.4	21(4)	27(4)	7(4)	8(5)	4(7)	22(4)	25(5)	23(4)
ELA.6.5	1(4)	8(1)	22(2)					
ELA.6.6	23(1)	6(1)						
IR.6.CS								
ELA.6.10	3(7)	2(5)	24(5)					
ELA.6.11	22(1)	21(1)	6(5)	7(1)	1(4)	26(6)		
ELA.6.12	6(1)	5(7)	23(1)	27(1)	28(5)			
IR.6.IKI								

ELA.6.15	29(5)						
ELA.6.16	30(6)	25(5)	26(5)				
ELA.6.17	27(1)	29(1)					
WL.6							
WL.6.TTP							
ELA.6.20							
ELA.6.21	40(18)	41(18)					
ELA.6.22							
WL.6.PDW							
ELA.6.23	41(15)						
ELA.6.24							
ELA.6.25							
WL.6.RBPK							
ELA.6.26							
ELA.6.27							
ELA.6.28	40(3)						
WL.6.RW							
ELA.6.29							
WL.6.CSE							
ELA.6.36	32(2)	31(10)	20(2)	18(5)	19(10)	39(12)	
ELA.6.37	33(5)	32(10)	20(10)	39(12)			
WL.6.KL							
ELA.6.38	40(9)	39(4)	41(15)				
WL.6.VAU							
ELA.6.39	2(2)						
ELA.6.40	16(6)	18(1)					
ELA.6.41	19(2)	20(2)	31(2)	33(1)	32(2)		

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

WV ELA 2019 Grade 6 B2 ELEM

Low		Medium		High	
3.6		10.8		18	
1	10199-4360	ELA.6.2:1	ELA.6.5:4	ELA.6.11:4	
2	10199-4362	ELA.6.10:5	ELA.6.39:2		
3	10199-4363	ELA.6.10:7			
4	10199-4364	ELA.6.4:7			
5	10199-4365	ELA.6.12:7			
6	10199-4367	ELA.6.6:1	ELA.6.11:5	ELA.6.12:1	
7	10199-4370	ELA.6.2:1	ELA.6.4:4	ELA.6.11:1	
8	10199-5273	ELA.6.4:5	ELA.6.5:1		
9	10199-11052	ELA.6.7:1	ELA.6.8:1	ELA.6.9:4	
10	10199-6963	ELA.6.1:2	ELA.6.2:6	ELA.6.8:1	
11	10199-6964	ELA.6.7:6			
12	10199-6966	ELA.6.1:6			
13	10199-6967	ELA.6.3:1	ELA.6.8:5		
14	10199-6971	ELA.6.1:4	ELA.6.7:1	ELA.6.8:1	
15	10199-6968	ELA.6.9:6			
16	10199-6972	ELA.6.40:6			
17	10199-6973	ELA.6.14:6			
18	REP10199-1356	ELA.6.36:5	ELA.6.40:1		
19	REP10199-1359	ELA.6.36:10	ELA.6.41:2		
20	REP10199-1363	ELA.6.36:2	ELA.6.37:10	ELA.6.41:2	
21	10199-7131	ELA.6.1:1	ELA.6.4:4	ELA.6.11:1	
22	10199-7135	ELA.6.4:4	ELA.6.5:2	ELA.6.11:1	
23	10199-7137	ELA.6.8:1	ELA.6.4:4	ELA.6.6:1	ELA.6.12:1
24	10199-7138	ELA.6.7:1	ELA.6.10:5		
25	10199-7141	ELA.6.4:5	ELA.6.16:5		
26	10199-7140	ELA.6.11:6	ELA.6.16:5		
27	10199-7146	ELA.6.4:4	ELA.6.12:1	ELA.6.17:1	
28	10199-7142	ELA.6.9:1	ELA.6.12:5		
29	10199-7143	ELA.6.15:5	ELA.6.17:1		

30 10199-7147	ELA.6.16:6		
31 10199-1337	ELA.6.36:10	ELA.6.41:2	
32 10199-1340	ELA.6.36:2	ELA.6.37:10	ELA.6.41:2
33 10199-1343	ELA.6.37:5	ELA.6.41:1	
34 REP10199-13964	ELA.6.8:6		
35 10199-13966	ELA.6.7:6		
36 REP10199-13970	ELA.6.2:4	ELA.6.3:1	ELA.6.8:2
37 REP10199-13982	ELA.6.2:1	ELA.6.3:5	
38 10199-14018	ELA.6.3:6		
39 CONVENTIONS10199-14166	ELA.6.36:12	ELA.6.37:12	ELA.6.38:4
40 ELABORATION10199-14166	ELA.6.21:18	ELA.6.28:3	ELA.6.38:9
41 ORGANIZATION10199-14166	ELA.6.21:18	ELA.6.23:15	ELA.6.38:15

Table 6.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 6 B2 ELEM

Low DOK		Matched DOK		High DOK

LR.6								
LR.6.KID								
ELA.6.1: [2]	10:(2)[2]	12:(6)[2]	14:(4)[2]	21:(1)[2]				
ELA.6.2: [2]	1:(1)[2]	7:(1)[2]	10:(6)[2]	36:(4)[2]	37:(1)[1]			
ELA.6.3: [2]	13:(1)[2]	36:(1)[2]	37:(5)[2]	38:(6)[2]				
LR.6.CS								
ELA.6.7: [2]	9:(1)[2]	11:(6)[2]	14:(1)[3]	24:(1)[1]	35:(6)[2]			
ELA.6.8: [2]	9:(1)[2]	10:(1)[2]	13:(5)[2]	14:(1)[2]	23:(1)[1]	34:(6)[2]	36:(2)[2]	
ELA.6.9: [2]	9:(4)[2]	15:(6)[2]	28:(1)[2]					
LR.6.IKI								
ELA.6.13								
ELA.6.14: [3]	17:(6)[2]							
IR.6								
IR.6.KID								
ELA.6.4: [2]	4:(7)[2]	7:(4)[2]	8:(5)[1]	21:(4)[2]	22:(4)[2]	23:(4)[2]	25:(5)[2]	27:(4)[2]
ELA.6.5: [2]	1:(4)[2]	8:(1)[2]	22:(2)[2]					
ELA.6.6: [2]	6:(1)[1]	23:(1)[2]						
IR.6.CS								
ELA.6.10: [2]	2:(5)[2]	3:(7)[2]	24:(5)[2]					
ELA.6.11: [2]	1:(4)[2]	6:(5)[2]	7:(1)[2]	21:(1)[2]	22:(1)[2]	26:(6)[2]		
ELA.6.12: [2]	5:(7)[2]	6:(1)[3]	23:(1)[2]	27:(1)[2]	28:(5)[2]			
IR.6.IKI								
ELA.6.15: [3]	29:(5)[2]							
ELA.6.16: [3]	25:(5)[2]	26:(5)[2]	30:(6)[2]					
ELA.6.17: [3]	27:(1)[2]	29:(1)[2]						
WL.6								

WL.6.TTP							
ELA.6.20							
ELA.6.21: [3]	40:(18)[3]	41:(18)[3]					
ELA.6.22							
WL.6.PDW							
ELA.6.23: [3]	41:(15)[3]						
ELA.6.24							
ELA.6.25							
WL.6.RBPK							
ELA.6.26							
ELA.6.27							
ELA.6.28: [3]	40:(3)[3]						
WL.6.RW							
ELA.6.29							
WL.6.CSE							
ELA.6.36: [1]	18:(5)[1]	19:(10)[1]	20:(2)[1]	31:(10)[1]	32:(2)[1]	39:(12)[1]	
ELA.6.37: [1]	20:(10)[1]	32:(10)[1]	33:(5)[1]	39:(12)[1]			
WL.6.KL							
ELA.6.38: [2]	39:(4)[1]	40:(9)[3]	41:(15)[3]				
WL.6.VAU							
ELA.6.39: [2]	2:(2)[1]						
ELA.6.40: [2]	16:(6)[2]	18:(1)[1]					
ELA.6.41: [2]	19:(2)[1]	20:(2)[1]	31:(2)[1]	32:(2)[1]	33:(1)[1]		

Table 6.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
CONF WV Grade 6B2 ELA ELEM GROUP

Low	Medium	High
3	9	15

1 10199-4360	Exact:6	
2 10199-4362	Exact:6	
3 10199-4363	Exact:6	
4 10199-4364	Partial:6	
5 10199-4365	Minimal:6	
6 10199-4367	Exact:6	
7 10199-4370	Exact:6	
8 10199-5273	Exact:6	
9 10199-11052	Exact:4	Partial:2
10 10199-6963	Exact:6	
11 10199-6964	Exact:5	Partial:1
12 10199-6966	Exact:6	
13 10199-6967	Exact:6	
14 10199-6971	Exact:6	
15 10199-6968	Exact:6	
16 10199-6972	Exact:5	Minimal:1
17 10199-6973	Exact:5	Partial:1
18 REP10199-1356	Exact:6	
19 REP10199-1359	Exact:12	
20 REP10199-1363	Exact:12	
21 10199-7131	Exact:6	
22 10199-7135	Exact:6	
23 10199-7137	Partial:6	
24 10199-7138	Exact:6	
25 10199-7141	Exact:5	Partial:1
26 10199-7140	Exact:6	
27 10199-7146	Exact:1	Partial:5
28 10199-7142	Exact:6	
29 10199-7143	Exact:6	
30 10199-7147	Exact:2	Partial:4

31 10199-1337	Exact:12	
32 10199-1340	Exact:10	Partial:2
33 10199-1343	Exact:6	
34 REP10199-13964	Exact:6	
35 10199-13966	Exact:6	
36 REP10199-13970	Exact:6	
37 REP10199-13982	Exact:5	Partial:1
38 10199-14018	Exact:5	Partial:1
39 CONVENTIONS10199-14166	Partial:2	Minimal:10
40 ELABORATION10199-14166	Exact:6	Partial:12
41 ORGANIZATION10199-14166	Exact:3	Partial:15

Grade 6 Batch 2 ELA West Virginia – Middle Grades Panel

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 6 B2 Number of Assessment Items - 41*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.6 Grade 6 Literary Reading	3	8	2 3	6 2	75 25	13.2	0.45	YES
IR.6 Grade 6 Informational Reading	3	9.8	2 3	6 3	66.67 33.33	18.4	0.55	YES
WL.6 Grade 6 Writing and Language	7	16	1 2 3 4	2 5 8 1	12.5 31.25 50 6.25	27.2	0.45	YES
Total	13	33.8	1 2 3 4	2 17 13 1	6 52 39 3	58.8	0.84	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 6 B2 Number of Assessment Items - 41*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.6 Grade 6 Literary Reading	3	8	13.2	0.45	6.04	3	92.53	5	1.43	3	YES
IR.6 Grade 6 Informational Reading	3	9.8	18.4	0.55	21.75	4	78.25	4	0	0	YES
WL.6 Grade 6 Writing and Language	7	16	27.2	0.45	2.22	2	71.19	17	26.59	16	YES
Total	13	33.8	58.8	0.84	9.18	1.5	78.23	8.9	12.59	7.7	

*Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Five Reviewers
WV ELA 2019 Grade 6 B2 Number of Assessment Items - 41*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.6 Grade 6 Literary Reading	3	8	13.2	0.45	7	0	87.5	0	YES	29	1	0.84	0.03	YES
IR.6 Grade 6 Informational Reading	3	9.8	18.4	0.55	7.4	0.89	75.56	8.75	YES	41	1	0.76	0.03	YES
WL.6 Grade 6 Writing and Language	7	16	27.2	0.45	6	0	37.5	0	NO	29	1	0.73	0.03	YES
Total	13	33.8	58.8	0.84	6.8	0.72	66.85	26		33	7	0.78	0.06	

*Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Five Reviewers WV ELA 2019 Grade 6 B2
Number of Assessment Items - 41*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.6 Grade 6 Literary Reading	YES	YES	YES	YES
IR.6 Grade 6 Informational Reading	YES	YES	YES	YES
WL.6 Grade 6 Writing and Language	YES	YES	NO	YES

Table 6.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 6 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1	2	2	2	2	2
2	2	2	2	2	2
3	2	2	2	2	2
4	2	2	2	2	2
5	2	2	2	2	2
6	2	2	2	2	2
7	2	2	2	2	2
8	2	2	2	2	2
9	2	2	2	2	2
10	2	2	2	2	2
11	2	2	2	2	2
12	2	2	2	2	2
13	2	2	2	2	2
14	2	2	2	2	2
15	2	2	2	2	2
16	1	2	1	2	1
17	2	2	2	2	3
18	2	1	2	1	1
19	2	1	2	1	1
20	1	2	2	1	1
21	2	2	2	2	2
22	2	2	2	2	2
23	2	2	2	2	2
24	2	2	2	2	2
25	2	2	2	2	2
26	2	2	2	2	2
27	2	2	2	2	2
28	2	2	2	2	2
29	2	2	2	2	2
30	2	2	2	2	2
31	2	1	2	1	1

32	1	1	2	1	1
33	1	1	2	1	1
34	2	2	2	2	2
35	2	2	3	2	2
36	2	2	2	2	2
37	2	2	2	2	2
38	2	2	2	2	2
39	1	2	1	1	1
40	3	3	3	3	3
41	3	3	3	3	3

Intraclass correlation - .9325

Pairwise Comparison - 0.88

Table 6.6

DOK Levels and Objectives Code by Each Reviewer

WV ELA 2019 Grade 6 B2

Number of Reviewers: Five

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11		
2	2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10		
3	2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10		
4	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4		
5	2	ELA.6.12			2	ELA.6.12			2	ELA.6.12			2	ELA.6.12			2	ELA.6.12		
6	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11		
7	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.6			2	ELA.6.4		
8	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4		
9	2	ELA.6.9			2	ELA.6.9			2	ELA.6.9			2	ELA.6.1			2	ELA.6.9		
10	2	ELA.6.2			2	ELA.6.2			2	ELA.6.2			2	ELA.6.1	ELA.6.2		2	ELA.6.2		
11	2	ELA.6.7			2	ELA.6.7			2	ELA.6.7			2	ELA.6.40			2	ELA.6.7		
12	2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.1		
13	2	ELA.6.3			2	ELA.6.8			2	ELA.6.8			2	ELA.6.8			2	ELA.6.8		
14	2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.8			2	ELA.6.1		
15	2	ELA.6.9			2	ELA.6.9			2	ELA.6.9			2	ELA.6.9			2	ELA.6.9		
16	1	ELA.6.40			2	ELA.6.40			1	ELA.6.40			2	ELA.6.40			1	ELA.6.40		
17	2	ELA.6.14			2	ELA.6.14			2	ELA.6.14			2	ELA.6.14			3	ELA.6.14		
18	2	ELA.6.36			1	ELA.6.36			2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
19	2	ELA.6.36			1	ELA.6.36			2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
20	1	ELA.6.37			2	ELA.6.37			2	ELA.6.37			1	ELA.6.37			1	ELA.6.37		
21	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.11		
22	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.5			2	ELA.6.4		
23	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.6			2	ELA.6.4		
24	2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10		
25	2	ELA.6.4	ELA.6.16		2	ELA.6.16			2	ELA.6.16			2	ELA.6.16			2	ELA.6.16		
26	2	ELA.6.16			2	ELA.6.16			2	ELA.6.16	ELA.6.11		2	ELA.6.11			2	ELA.6.16		
27	2	ELA.6.4			2	ELA.6.17			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4		
28	2	ELA.6.12			2	ELA.6.12			2	ELA.6.12			2	ELA.6.12			2	ELA.6.12		
29	2	ELA.6.15			2	ELA.6.15			2	ELA.6.15			2	ELA.6.15			2	ELA.6.15		
30	2	IR.6.IKI			2	ELA.6.17			2	IR.6.IKI			2	IR.6.IKI			2	IR.6.IKI		
31	2	ELA.6.36			1	ELA.6.36			2	ELA.6.36			1	ELA.6.37			1	ELA.6.36		

32	1	ELA.6.37		1	ELA.6.37		2	ELA.6.37		1	ELA.6.37		1	ELA.6.37		
33	1	ELA.6.37		1	ELA.6.37		2	ELA.6.37		1	ELA.6.37		1	ELA.6.37		
34	2	ELA.6.8		2	ELA.6.8		2	ELA.6.8		2	ELA.6.8		2	ELA.6.8		
35	2	ELA.6.7		2	ELA.6.7		3	ELA.6.7		2	ELA.6.7		2	ELA.6.7		
36	2	ELA.6.2		2	ELA.6.2		2	ELA.6.1	ELA.6.2	2	ELA.6.2		2	ELA.6.2		
37	2	ELA.6.2		2	ELA.6.2		2	ELA.6.2		2	ELA.6.3		2	ELA.6.2		
38	2	ELA.6.9		2	ELA.6.3		2	ELA.6.3		2	ELA.6.3		2	ELA.6.3		
39	1	ELA.6.36	ELA.6.37	2	ELA.6.36	ELA.6.37	1	ELA.6.36	ELA.6.37	1	ELA.6.36	ELA.6.37	1	ELA.6.36	ELA.6.37	
40	3	ELA.6.21		3	ELA.6.21		3	ELA.6.21		3	ELA.6.21		3	ELA.6.21		
41	3	ELA.6.21	ELA.6.23	ELA.6.38	3	ELA.6.21	ELA.6.38	ELA.6.23	3	ELA.6.23	ELA.6.21	ELA.6.38	3	ELA.6.21	ELA.6.38	ELA.6.23
Objective Pairwise Comparison: 0.81																
Standard Pairwise Comparison: 0.99																

Table 6.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV ELA 2019 Grade 6 B2

	Low		Medium			High		
	0		9				15	
LR.6								
LR.6.KID								
ELA.6.1	10(1)	12(5)	14(4)	9(1)	36(1)			
ELA.6.2	37(4)	10(5)	36(5)					
ELA.6.3	13(1)	37(1)	38(4)					
LR.6.CS								
ELA.6.7	35(5)	11(4)						
ELA.6.8	13(4)	14(1)	34(5)					
ELA.6.9	38(1)	15(5)	9(4)					
LR.6.IKI								
ELA.6.13								
ELA.6.14	17(5)							
IR.6								
IR.6.KID								
ELA.6.4	4(5)	7(4)	8(5)	21(4)	23(4)	22(4)	25(1)	27(4)
ELA.6.5	22(1)							
ELA.6.6	23(1)	7(1)						
IR.6.CS								
ELA.6.10	2(5)	3(5)	24(5)					
ELA.6.11	21(1)	1(5)	6(5)	26(2)				
ELA.6.12	5(5)	28(5)						
IR.6.IKI	30(4)							

ELA.6.15	29(5)						
ELA.6.16	26(4)	25(5)					
ELA.6.17	27(1)	30(1)					
WL.6							
WL.6.TTP							
ELA.6.20							
ELA.6.21	40(15)	41(15)					
ELA.6.22							
WL.6.PDW							
ELA.6.23	41(15)						
ELA.6.24							
ELA.6.25							
WL.6.RBPK							
ELA.6.26							
ELA.6.27							
ELA.6.28							
WL.6.RW							
ELA.6.29							
WL.6.CSE							
ELA.6.36	39(10)	31(8)	18(5)	19(10)			
ELA.6.37	20(10)	31(2)	32(10)	33(5)	39(10)		
WL.6.KL							
ELA.6.38	41(15)						
WL.6.VAU							
ELA.6.39							
ELA.6.40	16(5)	11(1)					
ELA.6.41							

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 6 B2

Low		Medium		High	
3		9		15	
1	10199-4360	ELA.6.11:5			
2	10199-4362	ELA.6.10:5			
3	10199-4363	ELA.6.10:5			
4	10199-4364	ELA.6.4:5			
5	10199-4365	ELA.6.12:5			
6	10199-4367	ELA.6.11:5			
7	10199-4370	ELA.6.4:4	ELA.6.6:1		
8	10199-5273	ELA.6.4:5			
9	10199-11052	ELA.6.1:1	ELA.6.9:4		
10	10199-6963	ELA.6.1:1	ELA.6.2:5		
11	10199-6964	ELA.6.7:4	ELA.6.40:1		
12	10199-6966	ELA.6.1:5			
13	10199-6967	ELA.6.3:1	ELA.6.8:4		
14	10199-6971	ELA.6.1:4	ELA.6.8:1		
15	10199-6968	ELA.6.9:5			
16	10199-6972	ELA.6.40:5			
17	10199-6973	ELA.6.14:5			
18	REP10199-1356	ELA.6.36:5			
19	REP10199-1359	ELA.6.36:10			
20	REP10199-1363	ELA.6.37:10			
21	10199-7131	ELA.6.4:4	ELA.6.11:1		
22	10199-7135	ELA.6.4:4	ELA.6.5:1		
23	10199-7137	ELA.6.4:4	ELA.6.6:1		
24	10199-7138	ELA.6.10:5			
25	10199-7141	ELA.6.4:1	ELA.6.16:5		
26	10199-7140	ELA.6.11:2	ELA.6.16:4		
27	10199-7146	ELA.6.4:4	ELA.6.17:1		
28	10199-7142	ELA.6.12:5			
29	10199-7143	ELA.6.15:5			
30	10199-7147	IR.6.IK1:4	ELA.6.17:1		

31 10199-1337	ELA.6.36:8	ELA.6.37:2	
32 10199-1340	ELA.6.37:10		
33 10199-1343	ELA.6.37:5		
34 REP10199-13964	ELA.6.8:5		
35 10199-13966	ELA.6.7:5		
36 REP10199-13970	ELA.6.1:1	ELA.6.2:5	
37 REP10199-13982	ELA.6.2:4	ELA.6.3:1	
38 10199-14018	ELA.6.3:4	ELA.6.9:1	
39 CONVENTIONS10199-14166	ELA.6.36:10	ELA.6.37:10	
40 ELABORATION10199-14166	ELA.6.21:15		
41 ORGANIZATION10199-14166	ELA.6.21:15	ELA.6.23:15	ELA.6.38:15

Table 6.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 6 B2

Low DOK		Matched DOK		High DOK

LR.6								
LR.6.KID								
ELA.6.1: [2]	9:(1)[2]	10:(1)[2]	12:(5)[2]	14:(4)[2]	36:(1)[2]			
ELA.6.2: [2]	10:(5)[2]	36:(5)[2]	37:(4)[2]					
ELA.6.3: [2]	13:(1)[2]	37:(1)[2]	38:(4)[2]					
LR.6.CS								
ELA.6.7: [2]	11:(4)[2]	35:(5)[2]						
ELA.6.8: [2]	13:(4)[2]	14:(1)[2]	34:(5)[2]					
ELA.6.9: [2]	9:(4)[2]	15:(5)[2]	38:(1)[2]					
LR.6.IKI								
ELA.6.13								
ELA.6.14: [3]	17:(5)[2]							
IR.6								
IR.6.KID								
ELA.6.4: [2]	4:(5)[2]	7:(4)[2]	8:(5)[2]	21:(4)[2]	22:(4)[2]	23:(4)[2]	25:(1)[2]	27:(4)[2]
ELA.6.5: [2]	22:(1)[2]							
ELA.6.6: [2]	7:(1)[2]	23:(1)[2]						
IR.6.CS								
ELA.6.10: [2]	2:(5)[2]	3:(5)[2]	24:(5)[2]					
ELA.6.11: [2]	1:(5)[2]	6:(5)[2]	21:(1)[2]	26:(2)[2]				
ELA.6.12: [2]	5:(5)[2]	28:(5)[2]						
IR.6.IKI: [3]	30:(4)[2]							
ELA.6.15: [3]	29:(5)[2]							
ELA.6.16: [3]	25:(5)[2]	26:(4)[2]						
ELA.6.17: [3]	27:(1)[2]	30:(1)[2]						
WL.6								
WL.6.TTP								
ELA.6.20								

ELA.6.21: [3]	40:(15)[3]	41:(15)[3]						
ELA.6.22								
WL.6.PDW								
ELA.6.23: [3]	41:(15)[3]							
ELA.6.24								
ELA.6.25								
WL.6.RBPK								
ELA.6.26								
ELA.6.27								
ELA.6.28								
WL.6.RW								
ELA.6.29								
WL.6.CSE								
ELA.6.36: [1]	18:(5)[1]	19:(10)[1]	31:(8)[2]	39:(10)[1]				
ELA.6.37: [1]	20:(10)[1]	31:(2)[1]	32:(10)[1]	33:(5)[1]	39:(10)[1]			
WL.6.KL								
ELA.6.38: [2]	41:(15)[3]							
WL.6.VAU								
ELA.6.39								
ELA.6.40: [2]	11:(1)[2]	16:(5)[1]						
ELA.6.41								

Table 6.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 CONF WV Grade 6B2 ELA

Low		Medium		High	
3		9		15	
1	10199-4360	Partial:6			
2	10199-4362	Negligible:6			
3	10199-4363	Exact:5		Partial:1	
4	10199-4364	Partial:6			
5	10199-4365	Minimal:4		Negligible:2	
6	10199-4367	Exact:4		Partial:1	
7	10199-4370	Minimal:6			
8	10199-5273	Exact:6			
9	10199-11052	Exact:4		Partial:2	
10	10199-6963	Exact:5		Partial:1	
11	10199-6964	Exact:5		Partial:1	
12	10199-6966	Exact:6			
13	10199-6967	Exact:6			
14	10199-6971	Exact:5		Partial:1	
15	10199-6968	Exact:6			
16	10199-6972	Minimal:6			
17	10199-6973	Exact:4		Partial:1	
18	REP10199-1356	Exact:6			
19	REP10199-1359	Exact:12			
20	REP10199-1363	Exact:12			
21	10199-7131	Exact:6			
22	10199-7135	Exact:2		Partial:4	
23	10199-7137	Partial:6			
24	10199-7138	Exact:6			
25	10199-7141	Partial:5		Negligible:1	
26	10199-7140	Exact:2		Partial:4	
27	10199-7146	Minimal:5		Negligible:1	
28	10199-7142	Exact:6			
29	10199-7143	Exact:6			

30 10199-7147	Minimal:6	
31 10199-1337	Exact:10	Minimal:2
32 10199-1340	Exact:10	Minimal:2
33 10199-1343	Exact:6	
34 REP10199-13964	Exact:6	
35 10199-13966	Exact:4	Partial:2
36 REP10199-13970	Exact:6	
37 REP10199-13982	Exact:1	Minimal:5
38 10199-14018	Exact:5	Partial:1
39 CONVENTIONS10199-14166	Exact:2	Minimal:10
40 ELABORATION10199-14166	Exact:9	Partial:9
41 ORGANIZATION10199-14166	Exact:3	Partial:15

Grade 6 Batch 3 ELA West Virginia

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 6 B3 Number of Assessment Items - 42*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.6 Grade 6 Literary Reading	3	8	2 3	6 2	75 25	15	1.22	YES
IR.6 Grade 6 Informational Reading	3	9	2 3	6 3	66.67 33.33	18.6	1.52	YES
WL.6 Grade 6 Writing and Language	7	16	1 2 3 4	2 5 8 1	12.5 31.25 50 6.25	27	1.22	YES
Total	13	33	1 2 3 4	2 17 13 1	6 52 39 3	60.6	1.82	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 6 B3 Number of Assessment Items - 42*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.6 Grade 6 Literary Reading	3	8	15	1.22	10.8	4	88.03	3	1.18	3	YES
IR.6 Grade 6 Informational Reading	3	9	18.6	1.52	6.36	4	93.64	4	0	0	YES
WL.6 Grade 6 Writing and Language	7	16	27	1.22	0.74	2	76.91	15	22.35	16	YES
Total	13	33	60.6	1.82	4.95	1.6	84.82	7.4	10.23	6.8	

*Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Five Reviewers
WV ELA 2019 Grade 6 B3 Number of Assessment Items - 42*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.6 Grade 6 Literary Reading	3	8	15	1.22	6	0.71	75	8.84	YES	32	3	0.81	0.04	YES
IR.6 Grade 6 Informational Reading	3	9	18.6	1.52	7.4	0.55	82.22	6.09	YES	40	2	0.82	0.06	YES
WL.6 Grade 6 Writing and Language	7	16	27	1.22	5.6	0.55	35	3.42	NO	28	2	0.69	0.06	WEAK
Total	13	33	60.6	1.82	6.3	0.95	64.07	25		33	6	0.77	0.06	

*Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Five Reviewers WV ELA 2019 Grade 6 B3
Number of Assessment Items - 42*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.6 Grade 6 Literary Reading	YES	YES	YES	YES
IR.6 Grade 6 Informational Reading	YES	YES	YES	YES
WL.6 Grade 6 Writing and Language	YES	YES	NO	WEAK

Table 6.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 6 B3

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1	2	1	2	1	1
2	2	1	2	1	1
3	1	1	2	1	1
4	2	2	2	2	2
5	2	2	2	2	2
6	2	2	2	2	2
7	2	1	2	2	2
8	2	2	2	2	2
9	2	2	2	2	2
10	2	2	2	2	2
11	2	2	2	2	2
12	2	2	2	2	2
13	2	2	2	2	2
14	2	2	2	2	2
15	2	2	2	2	2
16	2	2	2	2	2
17	2	2	2	2	2
18	2	2	2	2	2
19	2	2	2	2	2
20	2	2	2	2	2
21	2	2	2	2	2
22	3	3	2	2	2
23	2	2	2	2	2
24	2	2	2	2	2
25	2	2	2	2	2
26	2	2	2	2	2
27	2	2	2	2	2
28	2	2	2	2	2
29	2	2	2	2	2
30	2	2	2	2	2
31	2	2	2	2	2

32	2	2	2	2	2
33	2	2	2	2	2
34	3	2	2	2	3
35	2	3	2	2	3
36	2	2	2	2	2
37	2	1	2	1	1
38	1	1	2	1	1
39	1	1	2	1	1
40	1	2	1	3	1
41	3	3	3	3	3
42	3	3	3	3	3

Intraclass correlation - .9058

Pairwise Comparison - 0.86

Table 6.6

DOK Levels and Objectives Code by Each Reviewer

WV ELA 2019 Grade 6 B3

Number of Reviewers: Five

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.6.36			1	ELA.6.36			2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
2	2	ELA.6.36			1	ELA.6.36			2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
3	1	ELA.6.37			1	ELA.6.37			2	ELA.6.37			1	ELA.6.37			1	ELA.6.37		
4	2	ELA.6.7	ELA.6.1		2	ELA.6.1			2	ELA.6.1			2	ELA.6.8			2	ELA.6.1		
5	2	ELA.6.2			2	ELA.6.2			2	ELA.6.2			2	ELA.6.2			2	ELA.6.2		
6	2	ELA.6.3			2	ELA.6.3			2	ELA.6.3			2	ELA.6.3			2	ELA.6.3		
7	2	ELA.6.7			1	ELA.6.39			2	ELA.6.7	ELA.6.39		2	ELA.6.39	ELA.6.7		2	ELA.6.7		
8	2	ELA.6.7			2	ELA.6.7			2	ELA.6.7			2	ELA.6.7			2	ELA.6.7		
9	2	ELA.6.7			2	ELA.6.8			2	ELA.6.8			2	ELA.6.8			2	ELA.6.8		
10	2	ELA.6.6			2	ELA.6.6			2	ELA.6.6			2	ELA.6.1			2	ELA.6.6		
11	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.11		
12	2	ELA.6.3			2	ELA.6.16			2	ELA.6.4			2	ELA.6.6			2	ELA.6.4		
13	2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.39	ELA.6.10		2	ELA.6.10		
14	2	ELA.6.10			2	ELA.6.10			2	ELA.6.10			2	ELA.6.10	ELA.6.39		2	ELA.6.10		
15	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.11			2	ELA.6.4		
16	2	ELA.6.5			2	ELA.6.5			2	ELA.6.5	ELA.6.4		2	ELA.6.5			2	ELA.6.5		
17	2	ELA.6.12			2	ELA.6.12			2	ELA.6.12	ELA.6.4		2	ELA.6.12			2	ELA.6.12		
18	2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.11			2	ELA.6.1		
19	2	ELA.6.7			2	ELA.6.7			2	ELA.6.7			2	ELA.6.4			2	ELA.6.7		
20	2	ELA.6.1			2	ELA.6.9			2	ELA.6.1			2	ELA.6.1			2	ELA.6.1		
21	2	ELA.6.2			2	ELA.6.2			2	ELA.6.8			2	ELA.6.2			2	ELA.6.2		
22	3	ELA.6.2			3	ELA.6.14			2	ELA.6.14			2	ELA.6.14			2	ELA.6.14		
23	2	ELA.6.14			2	ELA.6.14			2	ELA.6.14			2	ELA.6.14			2	ELA.6.14		
24	2	ELA.6.2			2	ELA.6.8			2	ELA.6.2			2	ELA.6.1			2	ELA.6.2		
25	2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.1		
26	2	ELA.6.1			2	ELA.6.1			2	ELA.6.1			2	ELA.6.8			2	ELA.6.1		
27	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.4		
28	2	ELA.6.4			2	ELA.6.4			2	ELA.6.4			2	ELA.6.6			2	ELA.6.4		
29	2	ELA.6.5			2	ELA.6.5			2	ELA.6.6			2	ELA.6.4			2	ELA.6.5		
30	2	ELA.6.12			2	ELA.6.6			2	ELA.6.11			2	ELA.6.5			2	ELA.6.6		
31	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11			2	ELA.6.4			2	ELA.6.11		

32	2	ELA.6.11		2	ELA.6.11		2	ELA.6.11		2	ELA.6.11		2	ELA.6.11		
33	2	ELA.6.12		2	ELA.6.12		2	ELA.6.12		2	ELA.6.12		2	ELA.6.12		
34	3	ELA.6.17		2	ELA.6.17		2	ELA.6.12		2	ELA.6.17		3	ELA.6.17		
35	2	ELA.6.17		3	ELA.6.17		2	ELA.6.16	ELA.6.17	2	ELA.6.4		3	ELA.6.17		
36	2	ELA.6.10		2	ELA.6.10		2	ELA.6.10		2	ELA.6.11		2	ELA.6.10		
37	2	ELA.6.36		1	ELA.6.36		2	ELA.6.36		1	ELA.6.36		1	ELA.6.36		
38	1	ELA.6.37		1	ELA.6.37		2	ELA.6.37		1	ELA.6.37		1	ELA.6.37		
39	1	ELA.6.36		1	ELA.6.36		2	ELA.6.36		1	ELA.6.36	ELA.6.37	1	ELA.6.36		
40	1	ELA.6.36	ELA.6.37	2	ELA.6.36	ELA.6.37	1	ELA.6.37	ELA.6.36	3	ELA.6.21		1	ELA.6.37	ELA.6.36	
41	3	ELA.6.21		3	ELA.6.21		3	ELA.6.21		3	ELA.6.21		3	ELA.6.21		
42	3	ELA.6.21	ELA.6.23	ELA.6.38	3	ELA.6.21	ELA.6.23	ELA.6.38	3	ELA.6.21	ELA.6.23	ELA.6.38	3	ELA.6.21	ELA.6.23	ELA.6.38
Objective Pairwise Comparison: 0.69																
Standard Pairwise Comparison: 0.94																

Table 6.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV ELA 2019 Grade 6 B3

	Low			Medium			High			
	0			9			15			
LR.6										
LR.6.KID										
ELA.6.1	10(1)	18(4)	20(4)	25(5)	26(4)	24(1)	4(4)			
ELA.6.2	24(3)	21(4)	22(1)	5(5)						
ELA.6.3	6(5)	12(1)								
LR.6.CS										
ELA.6.7	4(1)	8(5)	9(1)	19(4)	7(4)					
ELA.6.8	26(1)	21(1)	24(1)	9(4)	4(1)					
ELA.6.9	20(1)									
LR.6.IKI										
ELA.6.13										
ELA.6.14	22(4)	23(5)								
IR.6										
IR.6.KID										
ELA.6.4	19(1)	27(5)	28(4)	29(1)	31(1)	12(2)	15(4)	35(1)	16(1)	17(1)
ELA.6.5	16(5)	29(3)	30(1)							
ELA.6.6	30(2)	29(1)	28(1)	12(1)	10(4)					
IR.6.CS										
ELA.6.10	14(5)	36(4)	13(5)							
ELA.6.11	36(1)	15(1)	18(1)	11(5)	30(1)	31(4)	32(5)			
ELA.6.12	33(5)	34(1)	30(1)	17(5)						
IR.6.IKI										

ELA.6.15										
ELA.6.16	12(1)	35(1)								
ELA.6.17	34(4)	35(4)								
WL.6										
WL.6.TTP										
ELA.6.20										
ELA.6.21	41(15)	42(15)	40(2)							
ELA.6.22										
WL.6.PDW										
ELA.6.23	42(15)									
ELA.6.24										
ELA.6.25										
WL.6.RBPk										
ELA.6.26										
ELA.6.27										
ELA.6.28	42(3)									
WL.6.RW										
ELA.6.29										
WL.6.CSE										
ELA.6.36	39(10)	37(5)	1(5)	2(10)	40(8)					
ELA.6.37	3(10)	38(10)	40(8)	39(2)						
WL.6.KL										
ELA.6.38	42(12)									
WL.6.VAU										
ELA.6.39	13(1)	7(3)	14(1)							
ELA.6.40										
ELA.6.41										

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 6 B3

Low		Medium		High	
3		9		15	
1 10199-1394	ELA.6.36:5				
2 10199-1408	ELA.6.36:10				
3 10199-1417	ELA.6.37:10				
4 10199-12113	ELA.6.1:4	ELA.6.7:1	ELA.6.8:1		
5 REP10199-12114	ELA.6.2:5				
6 REP10199-12115	ELA.6.3:5				
7 REP10199-12116	ELA.6.7:4	ELA.6.39:3			
8 REP10199-12118	ELA.6.7:5				
9 REP10199-12119	ELA.6.7:1	ELA.6.8:4			
10 REP10199-7330	ELA.6.1:1	ELA.6.6:4			
11 REP10199-7333	ELA.6.11:5				
12 REP10199-7335	ELA.6.3:1	ELA.6.4:2	ELA.6.6:1	ELA.6.16:1	
13 10199-7336	ELA.6.10:5	ELA.6.39:1			
14 10199-7339	ELA.6.10:5	ELA.6.39:1			
15 REP10199-12251	ELA.6.4:4	ELA.6.11:1			
16 REP10199-12254	ELA.6.4:1	ELA.6.5:5			
17 REP10199-13989	ELA.6.4:1	ELA.6.12:5			
18 REP10199-4907	ELA.6.1:4	ELA.6.11:1			
19 REP10199-4908	ELA.6.7:4	ELA.6.4:1			
20 REP10199-4912	ELA.6.1:4	ELA.6.9:1			
21 REP10199-4919	ELA.6.2:4	ELA.6.8:1			
22 REP10199-4914	ELA.6.2:1	ELA.6.14:4			
23 10199-4915	ELA.6.14:5				
24 REP10199-7931	ELA.6.1:1	ELA.6.2:3	ELA.6.8:1		
25 10199-6486	ELA.6.1:5				
26 REP10199-7919	ELA.6.1:4	ELA.6.8:1			
27 10199-7446	ELA.6.4:5				
28 10199-7447	ELA.6.4:4	ELA.6.6:1			
29 10199-7457	ELA.6.4:1	ELA.6.5:3	ELA.6.6:1		
30 10199-7459	ELA.6.5:1	ELA.6.6:2	ELA.6.11:1	ELA.6.12:1	

31 10199-7463	ELA.6.4:1	ELA.6.11:4	
32 10199-7464	ELA.6.11:5		
33 10199-7465	ELA.6.12:5		
34 10199-7478	ELA.6.12:1	ELA.6.17:4	
35 10199-7479	ELA.6.4:1	ELA.6.16:1	ELA.6.17:4
36 10199-7482	ELA.6.10:4	ELA.6.11:1	
37 REP10199-3285	ELA.6.36:5		
38 REP10199-3286	ELA.6.37:10		
39 REP10199-3287	ELA.6.36:10	ELA.6.37:2	
40 CONVENTIONS10199-14166	ELA.6.21:2	ELA.6.36:8	ELA.6.37:8
41 ELABORATION10199-14166	ELA.6.21:15		
42 ORGANIZATION10199-14166	ELA.6.21:15	ELA.6.23:15	ELA.6.28:3 ELA.6.38:12

Table 6.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 6 B3

Low DOK		Matched DOK		High DOK

LR.6										
LR.6.KID										
ELA.6.1: [2]	4:(4)[2]	10:(1)[2]	18:(4)[2]	20:(4)[2]	24:(1)[2]	25:(5)[2]	26:(4)[2]			
ELA.6.2: [2]	5:(5)[2]	21:(4)[2]	22:(1)[3]	24:(3)[2]						
ELA.6.3: [2]	6:(5)[2]	12:(1)[2]								
LR.6.CS										
ELA.6.7: [2]	4:(1)[2]	7:(4)[2]	8:(5)[2]	9:(1)[2]	19:(4)[2]					
ELA.6.8: [2]	4:(1)[2]	9:(4)[2]	21:(1)[2]	24:(1)[2]	26:(1)[2]					
ELA.6.9: [2]	20:(1)[2]									
LR.6.IKI										
ELA.6.13										
ELA.6.14: [3]	22:(4)[2]	23:(5)[2]								
IR.6										
IR.6.KID										
ELA.6.4: [2]	12:(2)[2]	15:(4)[2]	16:(1)[2]	17:(1)[2]	19:(1)[2]	27:(5)[2]	28:(4)[2]	29:(1)[2]	31:(1)[2]	35:(1)[2]
ELA.6.5: [2]	16:(5)[2]	29:(3)[2]	30:(1)[2]							
ELA.6.6: [2]	10:(4)[2]	12:(1)[2]	28:(1)[2]	29:(1)[2]	30:(2)[2]					
IR.6.CS										
ELA.6.10: [2]	13:(5)[2]	14:(5)[2]	36:(4)[2]							
ELA.6.11: [2]	11:(5)[2]	15:(1)[2]	18:(1)[2]	30:(1)[2]	31:(4)[2]	32:(5)[2]	36:(1)[2]			
ELA.6.12: [2]	17:(5)[2]	30:(1)[2]	33:(5)[2]	34:(1)[2]						
IR.6.IKI										
ELA.6.15										
ELA.6.16: [3]	12:(1)[2]	35:(1)[2]								
ELA.6.17: [3]	34:(4)[2]	35:(4)[2]								
WL.6										
WL.6.TTP										
ELA.6.20										

ELA.6.21: [3]	40:(2)[3]	41:(15)[3]	42:(15)[3]							
ELA.6.22										
WL.6.PDW										
ELA.6.23: [3]	42:(15)[3]									
ELA.6.24										
ELA.6.25										
WL.6.RBPK										
ELA.6.26										
ELA.6.27										
ELA.6.28: [3]	42:(3)[3]									
WL.6.RW										
ELA.6.29										
WL.6.CSE										
ELA.6.36: [1]	1:(5)[1]	2:(10)[1]	37:(5)[1]	39:(10)[1]	40:(8)[1]					
ELA.6.37: [1]	3:(10)[1]	38:(10)[1]	39:(2)[1]	40:(8)[1]						
WL.6.KL										
ELA.6.38: [2]	42:(12)[3]									
WL.6.VAU										
ELA.6.39: [2]	7:(3)[2]	13:(1)[2]	14:(1)[2]							
ELA.6.40										
ELA.6.41										

Table 6.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 CONF WV Grade 6B3 ELA

Low		Medium		High	
3		9		15	
1	10199-1394	Exact:6			
2	10199-1408	Exact:12			
3	10199-1417	Exact:12			
4	10199-12113	Exact:4	Partial:2		
5	REP10199-12114	Exact:6			
6	REP10199-12115	Exact:5	Partial:1		
7	REP10199-12116	Exact:4	Partial:2		
8	REP10199-12118	Exact:6			
9	REP10199-12119	Exact:5	Partial:1		
10	REP10199-7330	Exact:5	Partial:1		
11	REP10199-7333	Exact:6			
12	REP10199-7335	Exact:1	Partial:3	Minimal:1	Negligible:1
13	10199-7336	Exact:2	Partial:4		
14	10199-7339	Exact:5	Partial:1		
15	REP10199-12251	Exact:5	Minimal:1		
16	REP10199-12254	Exact:5	Partial:1		
17	REP10199-13989	Exact:6			
18	REP10199-4907	Exact:5	Partial:1		
19	REP10199-4908	Exact:5	Minimal:1		
20	REP10199-4912	Exact:4	Partial:2		
21	REP10199-4919	Exact:6			
22	REP10199-4914	Exact:6			
23	10199-4915	Exact:6			
24	REP10199-7931	Exact:3	Partial:3		
25	10199-6486	Exact:1	Partial:4	Minimal:1	
26	REP10199-7919	Exact:5	Partial:1		
27	10199-7446	Exact:6			
28	10199-7447	Exact:5	Partial:1		
29	10199-7457	Exact:5	Minimal:1		
30	10199-7459	Exact:3	Partial:3		

31 10199-7463	Exact:6		
32 10199-7464	Exact:6		
33 10199-7465	Exact:6		
34 10199-7478	Exact:6		
35 10199-7479	Exact:5	Partial:1	
36 10199-7482	Exact:6		
37 REP10199-3285	Exact:6		
38 REP10199-3286	Exact:12		
39 REP10199-3287	Exact:12		
40 CONVENTIONS10199-14166	Exact:2	Minimal:10	
41 ELABORATION10199-14166	Exact:15	Partial:3	
42 ORGANIZATION10199-14166	Exact:3	Partial:12	Negligible:3

Grade 6 Batch 4 ELA West Virginia

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Three Reviewers WV ELA 2019 Grade 6 B4 Number of Assessment Items - 42*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.6 Grade 6 Literary Reading	3	8	2 3	6 2	75 25	14.33	1.15	YES
IR.6 Grade 6 Informational Reading	3	9	2 3	6 3	66.67 33.33	18.67	2.08	YES
WL.6 Grade 6 Writing and Language	7	16	1 2 3 4	2 5 8 1	12.5 31.25 50 6.25	27.67	2.08	YES
Total	13	33	1 2 3 4	2 17 13 1	6 52 39 3	60.67	2.08	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Three Reviewers WV ELA 2019 Grade 6 B4 Number of Assessment Items - 42*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.6 Grade 6 Literary Reading	3	8	14.33	1.15	9.57	8	90.43	8	0	0	YES
IR.6 Grade 6 Informational Reading	3	9	18.67	2.08	8.31	10	91.69	10	0	0	YES
WL.6 Grade 6 Writing and Language	7	16	27.67	2.08	0	0	76.77	22	23.23	22	YES
Total	13	33	60.67	2.08	4.95	4.8	84.62	13.3	10.44	9	

*Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Three Reviewers
WV ELA 2019 Grade 6 B4 Number of Assessment Items - 42*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.6 Grade 6 Literary Reading	3	8	14.33	1.15	6	0	75	0	YES	31	3	0.79	0.02	YES
IR.6 Grade 6 Informational Reading	3	9	18.67	2.08	7.33	1.15	81.48	12.83	YES	40	3	0.75	0.11	YES
WL.6 Grade 6 Writing and Language	7	16	27.67	2.08	6	1	37.5	6.25	NO	29	5	0.68	0.02	WEAK
Total	13	33	60.67	2.08	6.4	0.77	64.66	24		33	6	0.74	0.06	

*Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Three Reviewers WV ELA 2019 Grade 6 B4
Number of Assessment Items - 42*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.6 Grade 6 Literary Reading	YES	YES	YES	YES
IR.6 Grade 6 Informational Reading	YES	YES	YES	YES
WL.6 Grade 6 Writing and Language	YES	YES	NO	WEAK

Table 6.5
Depth-of-Knowledge Levels by Item and Reviewers
Intraclass Correlation
WV ELA 2019 Grade 6 B4
Reviewer's DOK

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1			2	2	2
2			2	2	2
3			2	2	2
4			2	2	2
5			2	2	2
6			2	2	2
7			2	2	2
8			2	2	2
9			2	2	2
10			2	2	3
11			2	2	2
12			2	2	2
13			2	2	2
14			2	2	2
15			2	2	2
16			2	2	3
17			2	2	2
18			2	2	2
19			2	2	3
20			2	1	1
21			2	1	1
22			2	1	1
23			2	1	2
24			2	2	2
25			2	2	2
26			2	2	2
27			2	2	2
28			2	2	2
29			2	1	1

30			2	1	1
31			2	1	1
32			2	2	2
33			2	2	2
34			2	2	2
35			2	2	3
36			2	2	2
37			2	2	2
38			2	2	2
39			2	2	3
40			1	1	1
41			3	3	3
42			3	3	3

Intraclass correlation - .8703

Pairwise Comparison - 0.81

Table 6.6 DOK Levels and Objectives Code by Each Reviewer
WV ELA 2019 Grade 6 B4

Number of Reviewers: Three

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.6.5			2	ELA.6.2			2	ELA.6.5		
2	2	ELA.6.4			2	ELA.6.11			2	ELA.6.4		
3	2	ELA.6.10			2	ELA.6.11			2	ELA.6.11		
4	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11		
5	2	ELA.6.4			2	ELA.6.6			2	ELA.6.4		
6	2	ELA.6.10			2	ELA.6.39			2	ELA.6.10		
7	2	ELA.6.6			2	ELA.6.11			2	ELA.6.6		
8	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11		
9	2	ELA.6.12			2	ELA.6.12			2	ELA.6.12		
10	2	ELA.6.17			2	ELA.6.17			3	ELA.6.17		
11	2	ELA.6.7	ELA.6.40		2	ELA.6.40	ELA.6.7		2	ELA.6.7		
12	2	ELA.6.1			2	ELA.6.11			2	ELA.6.1		
13	2	ELA.6.7			2	ELA.6.4			2	ELA.6.7		
14	2	ELA.6.1			2	ELA.6.1			2	ELA.6.1		
15	2	ELA.6.8			2	ELA.6.1			2	ELA.6.2		
16	2	ELA.6.14			2	ELA.6.14			3	ELA.6.14		
17	2	ELA.6.2			2	ELA.6.8			2	ELA.6.2		
18	2	ELA.6.1			2	ELA.6.8			2	ELA.6.1		
19	2	ELA.6.14			2	ELA.6.14			3	ELA.6.14		
20	2	ELA.6.36			1	ELA.6.37			1	ELA.6.36		
21	2	ELA.6.37			1	ELA.6.37			1	ELA.6.37		
22	2	ELA.6.36			1	ELA.6.37			1	ELA.6.36		
23	2	ELA.6.2			1	ELA.6.37			2	ELA.6.2		
24	2	ELA.6.3			2	ELA.6.3			2	ELA.6.3		
25	2	ELA.6.7			2	ELA.6.39			2	ELA.6.7		
26	2	ELA.6.7			2	ELA.6.8			2	ELA.6.7		
27	2	ELA.6.7			2	ELA.6.7			2	ELA.6.7		
28	2	ELA.6.8			2	ELA.6.8			2	ELA.6.8		
29	2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
30	2	ELA.6.36			1	ELA.6.36			1	ELA.6.36		
31	2	ELA.6.37			1	ELA.6.37			1	ELA.6.37		
32	2	ELA.6.6			2	ELA.6.1			2	ELA.6.6		
33	2	ELA.6.11			2	ELA.6.11			2	ELA.6.11		
34	2	ELA.6.4			2	ELA.6.6			2	ELA.6.4		

35	2	ELA.6.15	ELA.6.17		2	ELA.6.12			3	ELA.6.17		
36	2	ELA.6.4			2	ELA.6.11			2	ELA.6.4		
37	2	ELA.6.4	ELA.6.5		2	ELA.6.5			2	ELA.6.5		
38	2	ELA.6.4	ELA.6.12		2	ELA.6.12			2	ELA.6.12		
39	2	ELA.6.15			2	ELA.6.11			3	ELA.6.15		
40	1	ELA.6.36	ELA.6.37		1	ELA.6.36	ELA.6.37		1	ELA.6.37	ELA.6.36	
41	3	ELA.6.21			3	ELA.6.21			3	ELA.6.21		
42	3	ELA.6.21	ELA.6.23	ELA.6.38	3	ELA.6.21	ELA.6.23	ELA.6.38	3	ELA.6.21	ELA.6.38	ELA.6.23
Objective Pairwise Comparison: 0.6												
Standard Pairwise Comparison: 0.88												

Table 6.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV ELA 2019 Grade 6 B4

	Low			Medium			High		
	0			5.4			9		
LR.6									
LR.6.KID									
ELA.6.1	12(2)	14(3)	15(1)	18(2)	32(1)				
ELA.6.2	17(2)	23(2)	15(1)	1(1)					
ELA.6.3	24(3)								
LR.6.CS									
ELA.6.7	25(2)	27(3)	13(2)	26(2)	11(3)				
ELA.6.8	26(1)	15(1)	17(1)	18(1)	28(3)				
ELA.6.9									
LR.6.IKI									
ELA.6.13									
ELA.6.14	19(3)	16(3)							
IR.6									
IR.6.KID									
ELA.6.4	5(2)	2(2)	13(1)	36(2)	34(2)	37(1)	38(1)		
ELA.6.5	1(2)	37(3)							
ELA.6.6	7(2)	5(1)	32(2)	34(1)					
IR.6.CS									
ELA.6.10	6(2)	3(1)							
ELA.6.11	3(2)	4(3)	2(1)	7(1)	8(3)	12(1)	33(3)	36(1)	39(1)
ELA.6.12	35(1)	9(3)	38(3)						
IR.6.IKI									
ELA.6.15	35(1)	39(2)							
ELA.6.16									
ELA.6.17	10(3)	35(2)							
WL.6									
WL.6.TTP									

ELA.6.20									
ELA.6.21	41(9)	42(9)							
ELA.6.22									
WL.6.PDW									
ELA.6.23	42(9)								
ELA.6.24									
ELA.6.25									
WL.6.RBPK									
ELA.6.26									
ELA.6.27									
ELA.6.28									
WL.6.RW									
ELA.6.29									
WL.6.CSE									
ELA.6.36	29(3)	30(6)	22(4)	20(2)	40(6)				
ELA.6.37	20(1)	21(6)	22(2)	23(1)	31(6)	40(6)			
WL.6.KL									
ELA.6.38	42(9)								
WL.6.VAU									
ELA.6.39	25(1)	6(1)							
ELA.6.40	11(2)								
ELA.6.41									

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 6 B4

Low		Medium		High	
1.8		5.4		9	
1	10199-8914	ELA.6.2:1	ELA.6.5:2		
2	10199-8927	ELA.6.4:2	ELA.6.11:1		
3	10199-8930	ELA.6.10:1	ELA.6.11:2		
4	10199-8932	ELA.6.11:3			
5	10199-8936	ELA.6.4:2	ELA.6.6:1		
6	10199-8940	ELA.6.10:2	ELA.6.39:1		
7	10199-12514	ELA.6.6:2	ELA.6.11:1		
8	10199-12515	ELA.6.11:3			
9	10199-12516	ELA.6.12:3			
10	10199-12517	ELA.6.17:3			
11	10199-4893	ELA.6.7:3	ELA.6.40:2		
12	REP10199-4907	ELA.6.1:2	ELA.6.11:1		
13	REP10199-4908	ELA.6.7:2	ELA.6.4:1		
14	REP10199-4912	ELA.6.1:3			
15	REP10199-4919	ELA.6.1:1	ELA.6.2:1	ELA.6.8:1	
16	REP10199-4914	ELA.6.14:3			
17	REP10199-7931	ELA.6.2:2	ELA.6.8:1		
18	REP10199-7919	ELA.6.1:2	ELA.6.8:1		
19	10199-7932	ELA.6.14:3			
20	REP10199-3285	ELA.6.36:2	ELA.6.37:1		
21	REP10199-3286	ELA.6.37:6			
22	REP10199-3287	ELA.6.36:4	ELA.6.37:2		
23	REP10199-12114	ELA.6.2:2	ELA.6.37:1		
24	REP10199-12115	ELA.6.3:3			
25	REP10199-12116	ELA.6.7:2	ELA.6.39:1		
26	10199-12117	ELA.6.7:2	ELA.6.8:1		
27	REP10199-12118	ELA.6.7:3			
28	REP10199-12119	ELA.6.8:3			
29	REP10199-1356	ELA.6.36:3			
30	REP10199-1359	ELA.6.36:6			
31	REP10199-1363	ELA.6.37:6			

32 REP10199-7330	ELA.6.1:1	ELA.6.6:2	
33 REP10199-7333	ELA.6.11:3		
34 REP10199-7335	ELA.6.4:2	ELA.6.6:1	
35 10199-12004	ELA.6.12:1	ELA.6.15:1	ELA.6.17:2
36 REP10199-12251	ELA.6.4:2	ELA.6.11:1	
37 REP10199-12254	ELA.6.4:1	ELA.6.5:3	
38 REP10199-13989	ELA.6.4:1	ELA.6.12:3	
39 10199-13990	ELA.6.11:1	ELA.6.15:2	
40 CONVENTIONS10199-14166	ELA.6.36:6	ELA.6.37:6	
41 ELABORATION10199-14166	ELA.6.21:9		
42 ORGANIZATION10199-14166	ELA.6.21:9	ELA.6.23:9	ELA.6.38:9

Table 6.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 6 B4

Low DOK		Matched DOK		High DOK

LR.6									
LR.6.KID									
ELA.6.1: [2]	12:(2)[2]	14:(3)[2]	15:(1)[2]	18:(2)[2]	32:(1)[2]				
ELA.6.2: [2]	1:(1)[2]	15:(1)[2]	17:(2)[2]	23:(2)[2]					
ELA.6.3: [2]	24:(3)[2]								
LR.6.CS									
ELA.6.7: [2]	11:(3)[2]	13:(2)[2]	25:(2)[2]	26:(2)[2]	27:(3)[2]				
ELA.6.8: [2]	15:(1)[2]	17:(1)[2]	18:(1)[2]	26:(1)[2]	28:(3)[2]				
ELA.6.9									
LR.6.IKI									
ELA.6.13									
ELA.6.14: [3]	16:(3)[2]	19:(3)[2]							
IR.6									
IR.6.KID									
ELA.6.4: [2]	2:(2)[2]	5:(2)[2]	13:(1)[2]	34:(2)[2]	36:(2)[2]	37:(1)[2]	38:(1)[2]		
ELA.6.5: [2]	1:(2)[2]	37:(3)[2]							
ELA.6.6: [2]	5:(1)[2]	7:(2)[2]	32:(2)[2]	34:(1)[2]					
IR.6.CS									
ELA.6.10: [2]	3:(1)[2]	6:(2)[2]							
ELA.6.11: [2]	2:(1)[2]	3:(2)[2]	4:(3)[2]	7:(1)[2]	8:(3)[2]	12:(1)[2]	33:(3)[2]	36:(1)[2]	39:(1)[2]
ELA.6.12: [2]	9:(3)[2]	35:(1)[2]	38:(3)[2]						
IR.6.IKI									
ELA.6.15: [3]	35:(1)[2]	39:(2)[2]							
ELA.6.16									
ELA.6.17: [3]	10:(3)[2]	35:(2)[2]							
WL.6									
WL.6.TTP									
ELA.6.20									
ELA.6.21: [3]	41:(9)[3]	42:(9)[3]							

ELA.6.22									
WL.6.PDW									
ELA.6.23: [3]	42:(9)[3]								
ELA.6.24									
ELA.6.25									
WL.6.RBPK									
ELA.6.26									
ELA.6.27									
ELA.6.28									
WL.6.RW									
ELA.6.29									
WL.6.CSE									
ELA.6.36: [1]	20:(2)[2]	22:(4)[2]	29:(3)[1]	30:(6)[1]	40:(6)[1]				
ELA.6.37: [1]	20:(1)[1]	21:(6)[1]	22:(2)[1]	23:(1)[1]	31:(6)[1]	40:(6)[1]			
WL.6.KL									
ELA.6.38: [2]	42:(9)[3]								
WL.6.VAU									
ELA.6.39: [2]	6:(1)[2]	25:(1)[2]							
ELA.6.40: [2]	11:(2)[2]								
ELA.6.41									

Grade 7 Batch 1 ELA West Virginia

Table 7.1 *Categorical Concurrence between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 7 B1 Number of Assessment Items - 43*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.7 Grade 7 Literary Reading	3	8	2 3	1 7	12.5 87.5	14	1.22	YES
IR.7 Grade 7 Informational Reading	3	9	2 3	1 8	11.11 88.89	17.4	1.14	YES
WL.7 Grade 7 Writing and Language	7	16	1 2 3 4	1 6 8 1	6.25 37.5 50 6.25	28.6	1.95	YES
Total	13	33	1 2 3 4	1 8 23 1	3 24 70 3	60	1.41	

Table 7.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 7 B1 Number of Assessment Items - 43*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.7 Grade 7 Literary Reading	3	8	14	1.22	62.05	10	37.95	10	0	0	NO
IR.7 Grade 7 Informational Reading	3	9	17.4	1.14	73.54	7	25.4	8	1.05	2	NO
WL.7 Grade 7 Writing and Language	7	16	28.6	1.95	15.18	11	68.73	8	16.09	5	YES
Total	13	33	60	1.41	43	8.8	49	7.8	8	2	

*Table 7.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Five Reviewers
WV ELA 2019 Grade 7 B1 Number of Assessment Items - 43*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.7 Grade 7 Literary Reading	3	8	14	1.22	6.6	0.55	82.5	6.85	YES	30	2	0.81	0.05	YES
IR.7 Grade 7 Informational Reading	3	9	17.4	1.14	7.8	0.84	86.67	9.3	YES	37	3	0.77	0.03	YES
WL.7 Grade 7 Writing and Language	7	16	28.6	1.95	6	1	37.5	6.25	NO	33	4	0.74	0.07	YES
Total	13	33	60	1.41	6.8	0.92	68.89	27		33	4	0.77	0.03	

*Table 7.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Five Reviewers WV ELA 2019 Grade 7 B1
Number of Assessment Items - 43*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.7 Grade 7 Literary Reading	YES	NO	YES	YES
IR.7 Grade 7 Informational Reading	YES	NO	YES	YES
WL.7 Grade 7 Writing and Language	YES	YES	NO	YES

Table 7.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 7 B1

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1	2	2	2	2	2
2	2	2	2	2	2
3	2	2		2	2
4	2	2	2	2	2
5	2	2	2	2	2
6	2	2	2	2	2
7	2	2	2	3	3
8	2	2	2	2	2
9	2	2	2	2	2
10	2	2	2	2	2
11	2	2	2	2	2
12	3	3	3	2	3
13	2	2	2	2	2
14	2	2	2	2	2
15	2	2	2	2	2
16	3	2	3	2	3
17	3	2	2	2	3
18	2	2	2	2	2
19	2	2	2	2	2
20	1	1	2	1	2
21	2	1	2	1	2
22	2	1	2	1	1
23	2	2	2	2	2
24	2	2	2	2	2
25	2	2	2	2	2
26	2	2	2	2	2
27	2	2	2	2	3
28	2	2	2	2	2
29	2	2	2	2	2
30	2	2	2	2	2
31	2	2	2	2	2

32	2	3	2	2	2
33	2	2	2	2	2
34	2	2	2	2	2
35	2	2	2	2	3
36	3	2	2	2	2
37	1	1	2	1	1
38	1	1	2	2	2
39	1	1	2	1	2
40	2	1	2	1	2
41	1	1	1	1	1
42	3	3	3	3	3
43	3	3	3	3	3

Intraclass correlation - .898

Pairwise Comparison - 0.82

Table 7.6
DOK Levels and Objectives Code by Each Reviewer
WV ELA 2019 Grade 7 B1

Number of Reviewers: Five

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4		
2	2	ELA.7.5			2	ELA.7.5			2	ELA.7.5			2	ELA.7.5			2	ELA.7.5		
3	2	ELA.7.10			2	ELA.7.10							2	ELA.7.39			2	ELA.7.10		
4	2	ELA.7.5			2	ELA.7.5			2	ELA.7.5			2	ELA.7.5			2	ELA.7.5		
5	2	ELA.7.5			2	ELA.7.5			2	ELA.7.5			2	ELA.7.17			2	ELA.7.5		
6	2	ELA.7.6			2	ELA.7.6			2	ELA.7.11			2	ELA.7.6			2	ELA.7.6		
7	2	ELA.7.6			2	ELA.7.6			2	ELA.7.4			3	ELA.7.23			3	ELA.7.6		
8	2	ELA.7.10			2	ELA.7.10			2	ELA.7.40	ELA.7.10		2	ELA.7.39			2	ELA.7.40		
9	2	ELA.7.12			2	ELA.7.12			2	ELA.7.12			2	ELA.7.12			2	ELA.7.12		
10	2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.11			2	ELA.7.4		
11	2	ELA.7.7			2	ELA.7.10			2	ELA.7.39	ELA.7.7		2	ELA.7.11			2	ELA.7.39		
12	3	ELA.7.14			3	ELA.7.14			3	ELA.7.14			2	ELA.7.39			3	ELA.7.14		
13	2	ELA.7.2			2	ELA.7.2			2	ELA.7.2			2	ELA.7.8			2	ELA.7.2		
14	2	ELA.7.7			2	ELA.7.7			2	ELA.7.1			2	ELA.7.7			2	ELA.7.7		
15	2	ELA.7.8			2	ELA.7.8			2	ELA.7.8			2	ELA.7.8			2	ELA.7.8		
16	3	ELA.7.14			2	ELA.7.14			3	ELA.7.14			2	ELA.7.3			3	ELA.7.14		
17	3	ELA.7.14			2	ELA.7.3			2	ELA.7.3			2	ELA.7.14			3	ELA.7.3		
18	2	ELA.7.7			2	ELA.7.7			2	ELA.7.7	ELA.7.39		2	ELA.7.7			2	ELA.7.7		
19	2	ELA.7.1			2	ELA.7.1			2	ELA.7.1			2	ELA.7.8			2	ELA.7.1		
20	1	ELA.7.37			1	ELA.7.36			2	ELA.7.36			1	ELA.7.36			2	ELA.7.36		
21	2	ELA.7.38			1	ELA.7.36			2	ELA.7.36			1	ELA.7.36			2	ELA.7.36		
22	2	ELA.7.37			1	ELA.7.37			2	ELA.7.37			1	ELA.7.37			1	ELA.7.37		
23	2	ELA.7.1			2	ELA.7.1			2	ELA.7.1			2	ELA.7.1			2	ELA.7.1		
24	2	ELA.7.1			2	ELA.7.1			2	ELA.7.1			2	ELA.7.1			2	ELA.7.3		
25	2	ELA.7.2			2	ELA.7.2			2	ELA.7.2			2	ELA.7.2			2	ELA.7.2		
26	2	ELA.7.1			2	ELA.7.3			2	ELA.7.1			2	ELA.7.1			2	ELA.7.3		
27	2	ELA.7.9			2	ELA.7.9			2	ELA.7.3			2	ELA.7.9			3	ELA.7.9		
28	2	ELA.7.7			2	ELA.7.7			2	ELA.7.7	ELA.7.39		2	ELA.7.39			2	ELA.7.7		
29	2	ELA.7.10			2	ELA.7.10			2	ELA.7.10			2	ELA.7.10			2	ELA.7.10		
30	2	ELA.7.11			2	ELA.7.11			2	ELA.7.11			2	ELA.7.5			2	ELA.7.11		

31	2	ELA.7.4			2	ELA.7.16			2	ELA.7.16			2	ELA.7.16			2	ELA.7.16		
32	2	ELA.7.4			3	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4		
33	2	ELA.7.11			2	ELA.7.11			2	ELA.7.11			2	ELA.7.11			2	ELA.7.11		
34	2	ELA.7.5			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4		
35	2	ELA.7.15			2	ELA.7.6			2	ELA.7.6			2	ELA.7.6			3	ELA.7.6		
36	3	ELA.7.15			2	ELA.7.15			2	ELA.7.15			2	ELA.7.15			2	ELA.7.5		
37	1	ELA.7.37			1	ELA.7.37			2	ELA.7.37			1	ELA.7.37			1	ELA.7.37		
38	1	ELA.7.37			1	ELA.7.37			2	ELA.7.37			2	ELA.7.37			2	ELA.7.37		
39	1	ELA.7.37			1	ELA.7.36			2	ELA.7.36			1	ELA.7.36			2	ELA.7.36		
40	2	ELA.7.36			1	ELA.7.36			2	ELA.7.36			1	ELA.7.36			2	ELA.7.36		
41	1	ELA.7.36	ELA.7.37		1	ELA.7.36	ELA.7.37		1	ELA.7.36	ELA.7.37		1	ELA.7.36	ELA.7.37		1	ELA.7.36	ELA.7.37	
42	3	ELA.7.21			3	ELA.7.21			3	ELA.7.21			3	ELA.7.21	ELA.7.38		3	ELA.7.21		
43	3	ELA.7.21	ELA.7.23	ELA.7.38	3	ELA.7.21	ELA.7.23	ELA.7.38	3	ELA.7.21	ELA.7.23	ELA.7.38	3	ELA.7.21	ELA.7.23		3	ELA.7.21	ELA.7.23	ELA.7.38
Objective Pairwise Comparison: 0.68																				
Standard Pairwise Comparison: 0.92																				

Table 7.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV ELA 2019 Grade 7 B1

	Low		Medium		High	
	0		9			15
LR.7						
LR.7.KID						
ELA.7.1	14(1)	19(4)	23(5)	24(4)	26(3)	
ELA.7.2	25(5)	13(4)				
ELA.7.3	16(1)	17(3)	24(1)	26(2)	27(1)	
LR.7.CS						
ELA.7.7	28(4)	18(5)	14(4)	11(2)		
ELA.7.8	13(1)	15(5)	19(1)			
ELA.7.9	27(4)					
LR.7.CS						
ELA.7.13						
ELA.7.14	17(2)	16(4)	12(4)			
IR.7						
IR.7.KID						
ELA.7.4	10(4)	1(5)	7(1)	31(1)	32(5)	34(4)
ELA.7.5	34(1)	36(1)	30(1)	2(5)	4(5)	5(4)
ELA.7.6	7(3)	6(4)	35(4)			
IR.7.CS						
ELA.7.10	3(3)	11(1)	29(5)	8(3)		
ELA.7.11	30(4)	33(5)	11(1)	10(1)	6(1)	
ELA.7.12	9(5)					
IR.7.IKI						

ELA.7.15	35(1)	36(4)					
ELA.7.16	31(4)						
ELA.7.17	5(1)						
WL.7							
WL.7.TTP							
ELA.7.20							
ELA.7.21	42(15)	43(15)					
ELA.7.22							
WL.7.PDW							
ELA.7.23	7(1)	43(15)					
ELA.7.24							
ELA.7.25							
WL.7.RBPK							
ELA.7.26							
ELA.7.27							
ELA.7.28							
WL.7.RW							
ELA.7.29							
WL.7.CSE							
ELA.7.36	20(4)	21(8)	40(5)	41(10)	39(8)		
ELA.7.37	39(2)	37(5)	38(5)	21(2)	20(1)	22(10)	41(10)
WL.7.KL							
ELA.7.38	21(2)	42(3)	43(12)				
WL.7.VAU							
ELA.7.39	8(1)	3(1)	11(2)	12(1)	28(2)	18(1)	
ELA.7.40	8(2)						
ELA.7.41							

Table 7.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
Grade 7B1 ELA WV 2019

Low		Medium		High	
4.8		14.4		24	
1	10199-5108	RL.7.1:5	RL.7.6:1		
2	10199-5111	RL.7.2:6			
3	10199-5112	RL.7.1:4	RL.7.3:2		
4	10199-5116	RL.7.1:2	RL.7.3:2	RL.7.6:2	
5	10199-5118	RL.7.1:1	RL.7.3:3	RL.7.6:2	
6	10199-5119	RL.7.4:5	L.7.4:1		
7	10199-5120	RL.7.4:6			
8	REP10199-5275	RI.7.1:5	RI.7.9:1		
9	REP10199-5277	RI.7.1:5	RI.7.6:1		
10	10199-5279	RI.7.2:6			
11	REP10199-5281	RL.7.7:1	RI.7.3:1	RI.7.7:2	RI.7.9:3
12	REP10199-5282	RI.7.4:6			
13	REP10199-5283	RI.7.6:6			
14	REP10199-5284	RI.7.5:6			
15	10199-5289	RI.7.4:6			
16	10199-5291	RI.7.5:5	RI.7.6:1		
17	REP10199-5295	RI.7.5:2	RI.7.9:4		
18	10199-4304	RL.7.2:3	RL.7.9:3		
19	10199-4307	RL.7.9:6			
20	10199-4308	RL.7.4:6			
21	10199-4310	RL.7.1:1	RL.7.5:2	RL.7.6:3	
22	10199-4312	RL.7.1:4	RL.7.3:2		
23	10199-4314	RL.7.2:6			
24	10199-4315	RL.7.4:3	RL.7.5:2	L.7.4:1	
25	10199-4318	RL.7.1:3	RL.7.2:1	RL.7.3:2	
26	10199-4320	RL.7.1:5	RL.7.3:1		
27	10199-4996	RL.7.2:1	RL.7.3:2	RL.7.5:2	RL.7.6:1
28	REP10199-9294	RI.7.1:3	RI.7.3:3		
29	10199-9295	RI.7.1:6			
30	REP10199-9296	RI.7.2:6			

31 REP10199-9298	RI.7.8:6		
32 REP10199-9304	RI.7.5:5	RI.7.6:1	
33 REP10199-12334	RI.7.4:5	RI.7.6:1	
34 REP10199-3150	L.7.1:3	L.7.2:3	
35 REP10199-3152	L.7.1:6	L.7.2:6	
36 REP10199-3155	RI.7.2:2	L.7.1:2	L.7.2:8
37 REP10199-3883	L.7.1:12	L.7.2:12	L.7.3:12
38 REP10199-3883	W.7.1:24		
39 REP10199-3883	W.7.1:24	W.7.4:24	

Table 7.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 7 B1

Low DOK		Matched DOK		High DOK

LR.7							
LR.7.KID							
ELA.7.1: [2]	14:(1)[2]	19:(4)[2]	23:(5)[2]	24:(4)[2]	26:(3)[2]		
ELA.7.2: [3]	13:(4)[2]	25:(5)[2]					
ELA.7.3: [3]	16:(1)[2]	17:(3)[2]	24:(1)[2]	26:(2)[2]	27:(1)[2]		
LR.7.CS							
ELA.7.7: [3]	11:(2)[2]	14:(4)[2]	18:(5)[2]	28:(4)[2]			
ELA.7.8: [3]	13:(1)[2]	15:(5)[2]	19:(1)[2]				
ELA.7.9: [3]	27:(4)[2]						
LR.7.IKI							
ELA.7.13							
ELA.7.14: [3]	12:(4)[3]	16:(4)[3]	17:(2)[2]				
IR.7							
IR.7.KID							
ELA.7.4: [2]	1:(5)[2]	7:(1)[2]	10:(4)[2]	31:(1)[2]	32:(5)[2]	34:(4)[2]	
ELA.7.5: [3]	2:(5)[2]	4:(5)[2]	5:(4)[2]	30:(1)[2]	34:(1)[2]	36:(1)[2]	
ELA.7.6: [3]	6:(4)[2]	7:(3)[2]	35:(4)[2]				
IR.7.CS							
ELA.7.10: [3]	3:(3)[2]	8:(3)[2]	11:(1)[2]	29:(5)[2]			
ELA.7.11: [3]	6:(1)[2]	10:(1)[2]	11:(1)[2]	30:(4)[2]	33:(5)[2]		
ELA.7.12: [3]	9:(5)[2]						
IR.7.IKI							
ELA.7.15: [3]	35:(1)[2]	36:(4)[2]					
ELA.7.16: [3]	31:(4)[2]						
ELA.7.17: [3]	5:(1)[2]						
WL.7							
WL.7.TTP							
ELA.7.20							

ELA.7.21: [3]	42:(15)[3]	43:(15)[3]					
ELA.7.22							
WL.7.PDW							
ELA.7.23: [3]	7:(1)[3]	43:(15)[3]					
ELA.7.24							
ELA.7.25							
WL.7.RBPK							
ELA.7.26							
ELA.7.27							
ELA.7.28							
WL.7.RW							
ELA.7.29							
WL.7.CSE							
ELA.7.36: [2]	20:(4)[2]	21:(8)[2]	39:(8)[2]	40:(5)[2]	41:(10)[1]		
ELA.7.37: [1]	20:(1)[1]	21:(2)[1]	22:(10)[1]	37:(5)[1]	38:(5)[2]	39:(2)[1]	41:(10)[1]
WL.7.KL							
ELA.7.38: [2]	21:(2)[2]	42:(3)[3]	43:(12)[3]				
WL.7.VAU							
ELA.7.39: [2]	3:(1)[2]	8:(1)[2]	11:(2)[2]	12:(1)[2]	18:(1)[2]	28:(2)[2]	
ELA.7.40: [2]	8:(2)[2]						
ELA.7.41							

Table 7.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
CONF WV Grade 7B1 ELA

Low		Medium		High	
3.6		10.8		18	
1 10199-5302	Exact:5	Partial:1			
2 10199-5320	Exact:5	Partial:1			
3 10199-5624	Exact:4	Minimal:2			
4 10199-8071	Exact:1	Partial:1	Minimal:3	Negligible:1	
5 10199-8073	Exact:4	Partial:1	Negligible:1		
6 10199-8075	Exact:3	Partial:3			
7 10199-8077	Exact:2	Partial:2	Minimal:1	Negligible:1	
8 10199-8078	Exact:4	Partial:1	Minimal:1		
9 10199-8079	Exact:3	Partial:2	Minimal:1		
10 10199-8080	Exact:4	Partial:2			
11 REP10199-5400	Exact:4	Partial:2			
12 REP10199-5401	Exact:6				
13 REP10199-5405	Exact:6				
14 REP10199-5408	Exact:5	Partial:1			
15 REP10199-5409	Exact:6				
16 10199-5411	Exact:4	Partial:2			
17 10199-5414	Exact:6				
18 REP10199-6589	Exact:6				
19 REP10199-6604	Exact:4	Partial:2			
20 10199-3150	Exact:6				
21 10199-3152	Exact:12				
22 10199-3155	Exact:12				
23 REP10199-5116	Partial:3	Minimal:3			
24 REP10199-5108	Exact:5	Partial:1			
25 REP10199-5111	Exact:6				
26 REP10199-5112	Exact:1	Partial:4	Minimal:1		
27 REP10199-5118	Exact:3	Partial:3			
28 10199-5119	Exact:3	Partial:2	Minimal:1		
29 10199-4325	Exact:5	Partial:1			
30 10199-4326	Exact:5	Partial:1			

31 10199-4328	Exact:4	Partial:2	
32 10199-4956	Exact:6		
33 10199-7522	Exact:3	Partial:3	
34 10199-7523	Exact:5	Partial:1	
35 10199-7524	Exact:2	Partial:4	
36 10199-7525	Partial:1	Minimal:2	Negligible:3
37 10199-3232	Exact:6		
38 10199-3233	Exact:6		
39 10199-3234	Exact:12		
40 10199-3235	Exact:6		
41 CONVENTIONS10199-14169	Exact:2	Minimal:8	Negligible:2
42 ELABORATION10199-14169	Exact:18		
43 ORGANIZATION10199-14169	Exact:3	Partial:12	Negligible:3

Grade 7 Batch 2 ELA West Virginia

Table 7.1 *Categorical Concurrence between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 7 B2*
Number of Assessment Items - 42

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.7 Grade 7 Literary Reading	3	8	2 3	1 7	12.5 87.5	13.4	0.55	YES
IR.7 Grade 7 Informational Reading	3	9	2 3	1 8	11.11 88.89	17.8	0.45	YES
WL.7 Grade 7 Writing and Language	7	16	1 2 3 4	1 6 8 1	6.25 37.5 50 6.25	27	1	YES
Total	13	33	1 2 3 4	1 8 23 1	3 24 70 3	58.2	0.45	

Table 7.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 7 B2*
Number of Assessment Items - 42

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.7 Grade 7 Literary Reading	3	8	13.4	0.55	67.14	16	32.86	16	0	0	NO
IR.7 Grade 7 Informational Reading	3	9	17.8	0.45	58.17	16	39.61	17	2.22	3	WEAK
WL.7 Grade 7 Writing and Language	7	16	27	1	16.48	13	71.14	9	12.39	9	YES
Total	13	33	58.2	0.45	40.89	13.9	52.58	12.3	6.53	3.7	

*Table 7.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Five Reviewers
WV ELA 2019 Grade 7 B2 Number of Assessment Items - 42*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.7 Grade 7 Literary Reading	3	8	13.4	0.55	5.6	0.55	70	6.85	YES	30	1	0.79	0.02	YES
IR.7 Grade 7 Informational Reading	3	9	17.8	0.45	6.6	0.55	73.33	6.09	YES	39	1	0.78	0.04	YES
WL.7 Grade 7 Writing and Language	7	16	27	1	6	1	37.5	6.25	NO	31	2	0.66	0.07	WEAK
Total	13	33	58.2	0.45	6.1	0.5	60.28	20		33	5	0.74	0.07	

*Table 7.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Five Reviewers WV ELA 2019 Grade 7 B2
Number of Assessment Items - 42*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.7 Grade 7 Literary Reading	YES	NO	YES	YES
IR.7 Grade 7 Informational Reading	YES	WEAK	YES	YES
WL.7 Grade 7 Writing and Language	YES	YES	NO	WEAK

Table 7.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 7 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1	1	1	2	1	2
2	1	1	2	1	2
3	1	1	2	1	1
4	2	2	2	2	2
5	2	2	2	2	2
6	2	2	2	2	2
7	2	2	2	2	2
8	2	2	2	2	3
9	2	2	2	2	3
10	2	2	2	2	2
11	2	2	2	2	3
12	2	2	2	2	2
13	2	2	2	2	2
14	2	2	2	2	2
15	2	2	2	2	3
16	2	2	2	2	3
17	1	1	1	1	1
18	2	2	2	2	3
19	3	2	2	2	3
20	2	2	2	2	3
21	2	2	2	2	2
22	2	2	2	2	2
23	2	2	2	2	2
24	3	3	2	2	3
25	2	2	2	2	2
26	2	2	2	2	2
27	2	2	2	2	3
28	2	2	2	2	2
29	3	3	2	2	3
30	3	2	2	2	3
31	1	1	2	1	1

32	1	1	2	1	2
33	1	1	2	1	2
34	1	1	2	1	2
35	2	2	2	2	2
36	2	3	2	2	2
37	2	2	2	2	2
38	2	2	2	2	3
39	2	2	2	2	2
40	1	1	1	1	1
41	3	3	3	3	3
42	3	3	3	3	3

Intraclass correlation - .9257

Pairwise Comparison - 0.76

Table 7.6
DOK Levels and Objectives Code by Each Reviewer
WV ELA 2019 Grade 7 B2

Number of Reviewers: Five

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	ELA.7.37			1	ELA.7.36			2	ELA.7.36			1	ELA.7.36			2	ELA.7.36		
2	1	ELA.7.37			1	ELA.7.36			2	ELA.7.36			1	ELA.7.36			2	ELA.7.36		
3	1	ELA.7.37			1	ELA.7.37			2	ELA.7.37			1	ELA.7.37			1	ELA.7.37		
4	2	ELA.7.5			2	ELA.7.5			2	ELA.7.5			2	ELA.7.5			2	ELA.7.5		
5	2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4		
6	2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.10			2	ELA.7.4		
7	2	ELA.7.10			2	ELA.7.10			2	ELA.7.10			2	ELA.7.10			2	ELA.7.10		
8	2	ELA.7.11			2	ELA.7.15			2	ELA.7.15			2	ELA.7.11			3	ELA.7.11		
9	2	ELA.7.12			2	ELA.7.12			2	ELA.7.12			2	ELA.7.11			3	ELA.7.12		
10	2	ELA.7.10			2	ELA.7.10			2	ELA.7.39			2	ELA.7.10			2	ELA.7.39		
11	2	ELA.7.11			2	ELA.7.11			2	ELA.7.11			2	ELA.7.11			3	ELA.7.11		
12	2	ELA.7.1			2	ELA.7.1			2	ELA.7.1			2	ELA.7.9			2	ELA.7.1		
13	2	ELA.7.2			2	ELA.7.2			2	ELA.7.2			2	ELA.7.2			2	ELA.7.2		
14	2	ELA.7.9			2	ELA.7.3			2	ELA.7.3			2	ELA.7.9			2	ELA.7.9		
15	2	ELA.7.7			2	ELA.7.7			2	ELA.7.1			2	ELA.7.40			3	ELA.7.7		
16	2	ELA.7.1			2	ELA.7.9			2	ELA.7.9			2	ELA.7.9			3	ELA.7.9		
17	1	ELA.7.1			1	ELA.7.1			1	ELA.7.1			1	ELA.7.1			1	ELA.7.1		
18	2	ELA.7.9			2	ELA.7.9			2	ELA.7.9			2	ELA.7.9			3	ELA.7.9		
19	3	ELA.7.14			2	ELA.7.14			2	ELA.7.14			2	ELA.7.14			3	ELA.7.14		
20	2	ELA.7.1			2	ELA.7.7			2	ELA.7.40			2	ELA.7.7			3	ELA.7.40		
21	2	ELA.7.5			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4		
22	2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4			2	ELA.7.4		
23	2	ELA.7.5			2	ELA.7.5			2	ELA.7.4			2	ELA.7.5			2	ELA.7.5		
24	3	ELA.7.4			3	ELA.7.4			2	ELA.7.4			2	ELA.7.15			3	ELA.7.15		
25	2	ELA.7.4			2	ELA.7.10			2	ELA.7.10			2	ELA.7.10			2	ELA.7.10		
26	2	ELA.7.4			2	ELA.7.12			2	ELA.7.4	ELA.7.12		2	ELA.7.4			2	ELA.7.12		
27	2	ELA.7.11			2	ELA.7.11			2	ELA.7.11			2	ELA.7.11			3	ELA.7.11		
28	2	ELA.7.4			2	ELA.7.11			2	ELA.7.11			2	ELA.7.11			2	ELA.7.11		
29	3	ELA.7.17			3	ELA.7.17			2	ELA.7.17			2	ELA.7.17			3	ELA.7.17		
30	3	ELA.7.17			2	ELA.7.17			2	ELA.7.17			2	ELA.7.17			3	ELA.7.17		
31	1	ELA.7.37			1	ELA.7.37			2	ELA.7.37			1	ELA.7.37			1	ELA.7.37		

32	1	ELA.7.37		1	ELA.7.36		2	ELA.7.36		1	ELA.7.36		2	ELA.7.36		
33	1	ELA.7.37		1	ELA.7.36		2	ELA.7.36		1	ELA.7.36		2	ELA.7.36		
34	1	ELA.7.37		1	ELA.7.37		2	ELA.7.37		1	ELA.7.37		2	ELA.7.36		
35	2	ELA.7.2		2	ELA.7.2		2	ELA.7.2		2	ELA.7.2		2	ELA.7.2		
36	2	ELA.7.1		3	ELA.7.2		2	ELA.7.1		2	ELA.7.1		2	ELA.7.3		
37	2	ELA.7.1		2	ELA.7.1		2	ELA.7.1		2	ELA.7.1		2	ELA.7.1		
38	2	ELA.7.9		2	ELA.7.9		2	ELA.7.3		2	ELA.7.9		3	ELA.7.9		
39	2	ELA.7.7		2	ELA.7.7		2	ELA.7.7		2	ELA.7.7		2	ELA.7.7		
40	1	ELA.7.36	ELA.7.37	1	ELA.7.36	ELA.7.37	1	ELA.7.37	ELA.7.36	1	ELA.7.36	ELA.7.37	1	ELA.7.36	ELA.7.37	
41	3	ELA.7.21		3	ELA.7.21		3	ELA.7.21		3	ELA.7.21	ELA.7.23	ELA.7.38	3	ELA.7.21	
42	3	ELA.7.21	ELA.7.23	ELA.7.38	3	ELA.7.21	ELA.7.23	ELA.7.38	3	ELA.7.21	ELA.7.23	ELA.7.38	3	ELA.7.21	ELA.7.23	ELA.7.38
Objective Pairwise Comparison: 0.7																
Standard Pairwise Comparison: 0.96																

Table 7.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV ELA 2019 Grade 7 B2

	Low			Medium			High		
	0			9				15	
LR.7									
LR.7.KID									
ELA.7.1	12(4)	15(1)	16(1)	17(5)	20(1)	37(5)	36(3)		
ELA.7.2	36(1)	35(5)	13(5)						
ELA.7.3	14(2)	38(1)	36(1)						
LR.7.CS									
ELA.7.7	39(5)	15(3)	20(2)						
ELA.7.8									
ELA.7.9	18(5)	16(4)	14(3)	12(1)	38(4)				
LR.7.CS									
ELA.7.13									
ELA.7.14	19(5)								
IR.7									
IR.7.KID									
ELA.7.4	21(4)	22(5)	23(1)	24(3)	25(1)	28(1)	26(3)	5(5)	6(4)
ELA.7.5	4(5)	23(4)	21(1)						
ELA.7.6									
IR.7.CS									
ELA.7.10	25(4)	6(1)	7(5)	10(3)					
ELA.7.11	9(1)	11(5)	8(3)	27(5)	28(4)				
ELA.7.12	9(4)	26(3)							
IR.7.IKI									
ELA.7.15	8(2)	24(2)							
ELA.7.16									

ELA.7.17	29(5)	30(5)							
WL.7									
WL.7.TTP									
ELA.7.20									
ELA.7.21	41(15)	42(15)							
ELA.7.22									
WL.7.PDW									
ELA.7.23	41(3)	42(12)							
ELA.7.24									
ELA.7.25									
WL.7.RBPK									
ELA.7.26									
ELA.7.27									
ELA.7.28	42(3)								
WL.7.RW									
ELA.7.29									
WL.7.CSE									
ELA.7.36	34(1)	32(8)	33(4)	1(4)	2(8)	40(10)			
ELA.7.37	2(2)	3(10)	1(1)	33(1)	32(2)	31(5)	34(4)	40(10)	
WL.7.KL									
ELA.7.38	41(3)	42(9)							
WL.7.VAU									
ELA.7.39	10(2)								
ELA.7.40	15(1)	20(2)							
ELA.7.41									

Table 7.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 7 B2

Low	Medium	High
3	9	15

1 REP10199-1420	ELA.7.36:4	ELA.7.37:1	
2 REP10199-1421	ELA.7.36:8	ELA.7.37:2	
3 REP10199-1424	ELA.7.37:10		
4 10199-13861	ELA.7.5:5		
5 10199-13868	ELA.7.4:5		
6 10199-13872	ELA.7.4:4	ELA.7.10:1	
7 10199-13873	ELA.7.10:5		
8 10199-13875	ELA.7.11:3	ELA.7.15:2	
9 10199-13892	ELA.7.11:1	ELA.7.12:4	
10 10199-13902	ELA.7.10:3	ELA.7.39:2	
11 10199-14022	ELA.7.11:5		
12 10199-8055	ELA.7.1:4	ELA.7.9:1	
13 10199-8056	ELA.7.2:5		
14 10199-8058	ELA.7.3:2	ELA.7.9:3	
15 10199-8059	ELA.7.1:1	ELA.7.7:3	ELA.7.40:1
16 10199-8060	ELA.7.1:1	ELA.7.9:4	
17 10199-8061	ELA.7.1:5		
18 10199-8062	ELA.7.9:5		
19 10199-8064	ELA.7.14:5		
20 10199-8065	ELA.7.1:1	ELA.7.7:2	ELA.7.40:2
21 10199-5275	ELA.7.4:4	ELA.7.5:1	
22 10199-5277	ELA.7.4:5		
23 10199-5279	ELA.7.4:1	ELA.7.5:4	
24 10199-5281	ELA.7.4:3	ELA.7.15:2	
25 10199-5282	ELA.7.4:1	ELA.7.10:4	
26 10199-5283	ELA.7.4:3	ELA.7.12:3	
27 10199-5284	ELA.7.11:5		
28 10199-5291	ELA.7.4:1	ELA.7.11:4	
29 10199-5295	ELA.7.17:5		
30 10199-5297	ELA.7.17:5		

31 REP10199-3240	ELA.7.37:5			
32 REP10199-3241	ELA.7.36:8	ELA.7.37:2		
33 REP10199-3242	ELA.7.36:4	ELA.7.37:1		
34 REP10199-3243	ELA.7.36:1	ELA.7.37:4		
35 REP10199-5111	ELA.7.2:5			
36 REP10199-5112	ELA.7.1:3	ELA.7.2:1	ELA.7.3:1	
37 10199-5113	ELA.7.1:5			
38 REP10199-5118	ELA.7.3:1	ELA.7.9:4		
39 REP10199-5120	ELA.7.7:5			
40 CONVENTIONS10199-14169	ELA.7.36:10	ELA.7.37:10		
41 ELABORATION10199-14169	ELA.7.21:15	ELA.7.23:3	ELA.7.38:3	
42 ORGANIZATION10199-14169	ELA.7.21:15	ELA.7.23:12	ELA.7.28:3	ELA.7.38:9

Table 7.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 7 B2

Low DOK		Matched DOK		High DOK

LR.7									
LR.7.KID									
ELA.7.1: [2]	12:(4)[2]	15:(1)[2]	16:(1)[2]	17:(5)[1]	20:(1)[2]	36:(3)[2]	37:(5)[2]		
ELA.7.2: [3]	13:(5)[2]	35:(5)[2]	36:(1)[3]						
ELA.7.3: [3]	14:(2)[2]	36:(1)[2]	38:(1)[2]						
LR.7.CS									
ELA.7.7: [3]	15:(3)[2]	20:(2)[2]	39:(5)[2]						
ELA.7.8									
ELA.7.9: [3]	12:(1)[2]	14:(3)[2]	16:(4)[2]	18:(5)[2]	38:(4)[2]				
LR.7.IKI									
ELA.7.13									
ELA.7.14: [3]	19:(5)[2]								
IR.7									
IR.7.KID									
ELA.7.4: [2]	5:(5)[2]	6:(4)[2]	21:(4)[2]	22:(5)[2]	23:(1)[2]	24:(3)[3]	25:(1)[2]	26:(3)[2]	28:(1)[2]
ELA.7.5: [3]	4:(5)[2]	21:(1)[2]	23:(4)[2]						
ELA.7.6									
IR.7.CS									
ELA.7.10: [3]	6:(1)[2]	7:(5)[2]	10:(3)[2]	25:(4)[2]					
ELA.7.11: [3]	8:(3)[2]	9:(1)[2]	11:(5)[2]	27:(5)[2]	28:(4)[2]				
ELA.7.12: [3]	9:(4)[2]	26:(3)[2]							
IR.7.IKI									
ELA.7.15: [3]	8:(2)[2]	24:(2)[2]							
ELA.7.16									
ELA.7.17: [3]	29:(5)[3]	30:(5)[2]							
WL.7									
WL.7.TTP									
ELA.7.20									

ELA.7.21: [3]	41:(15)[3]	42:(15)[3]							
ELA.7.22									
WL.7.PDW									
ELA.7.23: [3]	41:(3)[3]	42:(12)[3]							
ELA.7.24									
ELA.7.25									
WL.7.RBPK									
ELA.7.26									
ELA.7.27									
ELA.7.28: [3]	42:(3)[3]								
WL.7.RW									
ELA.7.29									
WL.7.CSE									
ELA.7.36: [2]	1:(4)[2]	2:(8)[2]	32:(8)[2]	33:(4)[2]	34:(1)[2]	40:(10)[1]			
ELA.7.37: [1]	1:(1)[1]	2:(2)[1]	3:(10)[1]	31:(5)[1]	32:(2)[1]	33:(1)[1]	34:(4)[1]	40:(10)[1]	
WL.7.KL									
ELA.7.38: [2]	41:(3)[3]	42:(9)[3]							
WL.7.VAU									
ELA.7.39: [2]	10:(2)[2]								
ELA.7.40: [2]	15:(1)[2]	20:(2)[2]							
ELA.7.41									

Table 7.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
CONF WV Grade 7B2 ELA

Low		Medium		High			
3		9		15			
1 REP10199-1420		Exact:6					
2 REP10199-1421		Exact:12					
3 REP10199-1424		Exact:12					
4 10199-13861		Exact:6					
5 10199-13868		Exact:6					
6 10199-13872		Exact:1		Partial:3		Minimal:2	
7 10199-13873		Exact:6					
8 10199-13875		Exact:1		Partial:4		Minimal:1	
9 10199-13892		Exact:4		Partial:2			
10 10199-13902		Exact:2		Partial:3		Negligible:1	
11 10199-14022		Exact:5		Partial:1			
12 10199-8055		Exact:6					
13 10199-8056		Exact:5		Partial:1			
14 10199-8058		Exact:1		Partial:4		Minimal:1	
15 10199-8059		Exact:4		Minimal:2			
16 10199-8060		Exact:6					
17 10199-8061		Partial:3		Minimal:2		Negligible:1	
18 10199-8062		Exact:5		Partial:1			
19 10199-8064		Exact:4		Partial:2			
20 10199-8065		Exact:2		Partial:2		Minimal:2	
21 10199-5275		Exact:4		Partial:2			
22 10199-5277		Exact:6					
23 10199-5279		Exact:5		Partial:1			
24 10199-5281		Partial:2		Minimal:4			
25 10199-5282		Exact:6					
26 10199-5283		Exact:4		Partial:1		Negligible:1	
27 10199-5284		Exact:5		Minimal:1			
28 10199-5291		Exact:4		Partial:2			
29 10199-5295		Exact:4		Partial:1		Minimal:1	
30 10199-5297		Exact:4		Partial:1		Minimal:1	

31 REP10199-3240	Exact:6		
32 REP10199-3241	Exact:12		
33 REP10199-3242	Exact:6		
34 REP10199-3243	Exact:6		
35 REP10199-5111	Exact:6		
36 REP10199-5112	Exact:1	Partial:4	Minimal:1
37 10199-5113	Exact:6		
38 REP10199-5118	Exact:3	Partial:2	Minimal:1
39 REP10199-5120	Exact:6		
40 CONVENTIONS10199-14169	Exact:2	Minimal:10	
41 ELABORATION10199-14169	Exact:15	Minimal:3	
42 ORGANIZATION10199-14169	Exact:3	Partial:12	Minimal:3

Grade 7 Batch 3 ELA West Virginia

Table 7.1 *Categorical Concurrence between Standards and Assessment as Rated by Three Reviewers WV ELA 2019 Grade 7 B3 Number of Assessment Items - 41*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.7 Grade 7 Literary Reading	3	8	2 3	1 7	12.5 87.5	13.33	0.58	YES
IR.7 Grade 7 Informational Reading	3	9	2 3	1 8	11.11 88.89	16.33	1.15	YES
WL.7 Grade 7 Writing and Language	7	16	1 2 3 4	1 6 8 1	6.25 37.5 50 6.25	25.33	2.89	YES
Total	13	33	1 2 3 4	1 8 23 1	3 24 70 3	54.99	3.46	

Table 7.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Three Reviewers WV ELA 2019 Grade 7 B3 Number of Assessment Items - 41*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.7 Grade 7 Literary Reading	3	8	13.33	0.58	67.03	20	32.97	20	0	0	NO
IR.7 Grade 7 Informational Reading	3	9	16.33	1.15	70.07	15	29.93	15	0	0	NO
WL.7 Grade 7 Writing and Language	7	16	25.33	2.89	25.7	13	66.89	10	7.41	6	YES
Total	13	33	54.99	3.46	48.48	15.7	47.88	13.1	3.64	3	

Table 7.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Three Reviewers
WV ELA 2019 Grade 7 B3 Number of Assessment Items - 41

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.7 Grade 7 Literary Reading	3	8	13.33	0.58	6.67	0.58	83.33	7.22	YES	31	2	0.81	0.05	YES
IR.7 Grade 7 Informational Reading	3	9	16.33	1.15	7.33	1.15	81.48	12.83	YES	38	2	0.82	0.03	YES
WL.7 Grade 7 Writing and Language	7	16	25.33	2.89	5.67	0.58	35.42	3.61	NO	32	0	0.74	0.03	YES
Total	13	33	54.99	3.46	6.6	0.84	66.74	27		34	4	0.79	0.04	

Table 7.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Three Reviewers
WV ELA 2019 Grade 7 B3 Number of Assessment Items - 41

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.7 Grade 7 Literary Reading	YES	NO	YES	YES
IR.7 Grade 7 Informational Reading	YES	NO	YES	YES
WL.7 Grade 7 Writing and Language	YES	YES	NO	YES

Table 7.5 *Depth-of-Knowledge Levels by Item and Reviewers*
Intraclass Correlation WV ELA 2019 Grade 7 B3

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1		1		1	2
2		1		1	2
3		1		1	1
4		2		2	2
5		2		2	2
6		2		2	3
7		2		2	2
8		2		2	2
9		2		2	2
10		2		2	2
11		2		2	2
12		2		2	1
13		3		2	2
14		2		2	2
15		3		2	3
16		2		2	2
17		2		2	2
18		2		2	3
19		2		2	2
20		2		2	2
21		2		2	2
22		2		2	2
23		2		2	3
24		2		2	3
25		2		2	2
26		2		2	2
27		2		2	2
28		2		2	2
29		2		2	2
30		1		1	1
31		1		1	2

32		1		1	2
33		1		1	2
34		2		2	2
35		2		2	2
36		2		2	2
37		3		2	3
38		2		2	3
39		1		1	1
40		3		3	3
41		3		3	3

Intraclass correlation - .9026

Pairwise Comparison - 0.77

Table 7.6
DOK Levels and Objectives Code by Each Reviewer
WV ELA 2019 Grade 7 B3

Number of Reviewers: Three

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	ELA.7.36			1	ELA.7.36			2	ELA.7.36		
2	1	ELA.7.36			1	ELA.7.36			2	ELA.7.36		
3	1	ELA.7.37			1	ELA.7.37			1	ELA.7.37		
4	2	ELA.7.4			2	ELA.7.12			2	ELA.7.4		
5	2	ELA.7.4			2	ELA.7.4			2	ELA.7.4		
6	2	ELA.7.4			2	ELA.7.4			3	ELA.7.6		
7	2	ELA.7.10			2	ELA.7.12			2	ELA.7.10		
8	2	ELA.7.11			2	ELA.7.11			2	ELA.7.11		
9	2	ELA.7.12			2	ELA.7.12			2	ELA.7.12		
10	2	ELA.7.12			2	ELA.7.12			2	ELA.7.12		
11	2	ELA.7.5			2	ELA.7.17			2	ELA.7.17		
12	2	ELA.7.39			2	ELA.7.39			1	ELA.7.39		
13	3	ELA.7.5			2	ELA.7.5			2	ELA.7.5		
14	2	ELA.7.10			2	ELA.7.39			2	ELA.7.10		
15	3	ELA.7.14			2	ELA.7.14			3	ELA.7.14		
16	2	ELA.7.1			2	ELA.7.1			2	ELA.7.1		
17	2	ELA.7.2			2	ELA.7.8			2	ELA.7.2		
18	2	ELA.7.3			2	ELA.7.8			3	ELA.7.3		
19	2	ELA.7.7			2	ELA.7.7			2	ELA.7.7		
20	2	ELA.7.8			2	ELA.7.8			2	ELA.7.8		
21	2	ELA.7.7			2	ELA.7.7			2	ELA.7.7		
22	2	ELA.7.1			2	ELA.7.8			2	ELA.7.1		
23	2	ELA.7.11			2	ELA.7.8			3	ELA.7.11		
24	2	ELA.7.16			2	ELA.7.16			3	ELA.7.16		
25	2	ELA.7.11			2	ELA.7.11			2	ELA.7.11		
26	2	ELA.7.4			2	ELA.7.6			2	ELA.7.4		
27	2	ELA.7.4			2	ELA.7.6			2	ELA.7.4		
28	2	ELA.7.10			2	ELA.7.10			2	ELA.7.10		
29	2	ELA.7.5			2	ELA.7.5			2	ELA.7.5		
30	1	ELA.7.37			1	ELA.7.37			1	ELA.7.37		
31	1	ELA.7.36			1	ELA.7.36			2	ELA.7.36		

32	1	ELA.7.36			1	ELA.7.36			2	ELA.7.36		
33	1	ELA.7.37			1	ELA.7.37			2	ELA.7.36		
34	2	ELA.7.3			2	ELA.7.3			2	ELA.7.3		
35	2	ELA.7.8			2	ELA.7.3			2	ELA.7.8		
36	2	ELA.7.9			2	ELA.7.1			2	ELA.7.1		
37	3	ELA.7.7			2	ELA.7.7			3	ELA.7.7		
38	2	ELA.7.9			2	ELA.7.9			3	ELA.7.9		
39	1	ELA.7.36	ELA.7.37		1	ELA.7.36	ELA.7.37		1	ELA.7.36	ELA.7.37	
40	3	ELA.7.20			3	ELA.7.20			3	ELA.7.20		
41	3	ELA.7.20	ELA.7.23	ELA.7.38	3	ELA.7.23			3	ELA.7.38	ELA.7.20	ELA.7.23
Objective Pairwise Comparison: 0.74												
Standard Pairwise Comparison: 0.97												

Table 7.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV ELA 2019 Grade 7 B3

	Low		Medium		High	
	0		5.4			9
LR.7						
LR.7.KID						
ELA.7.1	16(3)	22(2)	36(2)			
ELA.7.2	17(2)					
ELA.7.3	18(2)	34(3)	35(1)			
LR.7.CS						
ELA.7.7	37(3)	21(3)	19(3)			
ELA.7.8	20(3)	18(1)	17(1)	22(1)	23(1)	35(2)
ELA.7.9	38(3)	36(1)				
LR.7.IKI						
ELA.7.13						
ELA.7.14	15(3)					
IR.7						
IR.7.KID						
ELA.7.4	4(2)	26(2)	27(2)	5(3)	6(2)	
ELA.7.5	11(1)	13(3)	29(3)			
ELA.7.6	6(1)	27(1)	26(1)			
IR.7.CS						
ELA.7.10	28(3)	7(2)	14(2)			
ELA.7.11	8(3)	25(3)	23(2)			
ELA.7.12	4(1)	9(3)	10(3)	7(1)		
IR.7.IKI						

ELA.7.15						
ELA.7.16	24(3)					
ELA.7.17	11(2)					
WL.7						
WL.7.TTP						
ELA.7.20	40(9)	41(6)				
ELA.7.21						
ELA.7.22						
WL.7.PDW						
ELA.7.23	41(9)					
ELA.7.24						
ELA.7.25						
WL.7.RBPK						
ELA.7.26						
ELA.7.27						
ELA.7.28						
WL.7.RW						
ELA.7.29						
WL.7.CSE						
ELA.7.36	39(6)	31(6)	32(3)	33(1)	1(3)	2(6)
ELA.7.37	3(6)	33(2)	30(3)	39(6)		
WL.7.KL						
ELA.7.38	41(6)					
WL.7.VAU						
ELA.7.39	12(3)	14(1)				
ELA.7.40						
ELA.7.41						

Table 7.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 7 B3

Low		Medium		High	
1.8		5.4		9	
1	10199-1406	ELA.7.36:3			
2	10199-1407	ELA.7.36:6			
3	10199-1409	ELA.7.37:6			
4	10199-5226	ELA.7.4:2	ELA.7.12:1		
5	10199-5230	ELA.7.4:3			
6	10199-5231	ELA.7.4:2	ELA.7.6:1		
7	10199-5233	ELA.7.10:2	ELA.7.12:1		
8	10199-5234	ELA.7.11:3			
9	10199-5235	ELA.7.12:3			
10	10199-5236	ELA.7.12:3			
11	10199-5238	ELA.7.5:1	ELA.7.17:2		
12	10199-5625	ELA.7.39:3			
13	10199-5831	ELA.7.5:3			
14	REP10199-5400	ELA.7.10:2	ELA.7.39:1		
15	REP10199-5401	ELA.7.14:3			
16	10199-5403	ELA.7.1:3			
17	REP10199-5405	ELA.7.2:2	ELA.7.8:1		
18	10199-5407	ELA.7.3:2	ELA.7.8:1		
19	REP10199-5408	ELA.7.7:3			
20	REP10199-5409	ELA.7.8:3			
21	REP10199-6589	ELA.7.7:3			
22	REP10199-6604	ELA.7.1:2	ELA.7.8:1		
23	10199-7515	ELA.7.8:1	ELA.7.11:2		
24	REP10199-7518	ELA.7.16:3			
25	REP10199-7516	ELA.7.11:3			
26	REP10199-7510	ELA.7.4:2	ELA.7.6:1		
27	REP10199-7511	ELA.7.4:2	ELA.7.6:1		
28	REP10199-7514	ELA.7.10:3			
29	REP10199-7512	ELA.7.5:3			
30	REP10199-3240	ELA.7.37:3			

31 REP10199-3241	ELA.7.36:6		
32 REP10199-3242	ELA.7.36:3		
33 REP10199-3243	ELA.7.36:1	ELA.7.37:2	
34 10199-3910	ELA.7.3:3		
35 10199-3913	ELA.7.3:1	ELA.7.8:2	
36 10199-3911	ELA.7.1:2	ELA.7.9:1	
37 10199-4834	ELA.7.7:3		
38 10199-4837	ELA.7.9:3		
39 CONVENTIONS10199-14168	ELA.7.36:6	ELA.7.37:6	
40 ELABORATION10199-14168	ELA.7.20:9		
41 ORGANIZATION10199-14168	ELA.7.20:6	ELA.7.23:9	ELA.7.38:6

Table 7.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 7 B3

Low DOK		Matched DOK		High DOK

LR.7						
LR.7.KID						
ELA.7.1: [2]	16:(3)[2]	22:(2)[2]	36:(2)[2]			
ELA.7.2: [3]	17:(2)[2]					
ELA.7.3: [3]	18:(2)[2]	34:(3)[2]	35:(1)[2]			
LR.7.CS						
ELA.7.7: [3]	19:(3)[2]	21:(3)[2]	37:(3)[3]			
ELA.7.8: [3]	17:(1)[2]	18:(1)[2]	20:(3)[2]	22:(1)[2]	23:(1)[2]	35:(2)[2]
ELA.7.9: [3]	36:(1)[2]	38:(3)[2]				
LR.7.IKI						
ELA.7.13						
ELA.7.14: [3]	15:(3)[3]					
IR.7						
IR.7.KID						
ELA.7.4: [2]	4:(2)[2]	5:(3)[2]	6:(2)[2]	26:(2)[2]	27:(2)[2]	
ELA.7.5: [3]	11:(1)[2]	13:(3)[2]	29:(3)[2]			
ELA.7.6: [3]	6:(1)[3]	26:(1)[2]	27:(1)[2]			
IR.7.CS						
ELA.7.10: [3]	7:(2)[2]	14:(2)[2]	28:(3)[2]			
ELA.7.11: [3]	8:(3)[2]	23:(2)[2]	25:(3)[2]			
ELA.7.12: [3]	4:(1)[2]	7:(1)[2]	9:(3)[2]	10:(3)[2]		
IR.7.IKI						
ELA.7.15						
ELA.7.16: [3]	24:(3)[2]					
ELA.7.17: [3]	11:(2)[2]					
WL.7						
WL.7.TTP						
ELA.7.20: [3]	40:(9)[3]	41:(6)[3]				

ELA.7.21						
ELA.7.22						
WL.7.PDW						
ELA.7.23: [3]	41:(9)[3]					
ELA.7.24						
ELA.7.25						
WL.7.RBPK						
ELA.7.26						
ELA.7.27						
ELA.7.28						
WL.7.RW						
ELA.7.29						
WL.7.CSE						
ELA.7.36: [2]	1:(3)[1]	2:(6)[1]	31:(6)[1]	32:(3)[1]	33:(1)[2]	39:(6)[1]
ELA.7.37: [1]	3:(6)[1]	30:(3)[1]	33:(2)[1]	39:(6)[1]		
WL.7.KL						
ELA.7.38: [2]	41:(6)[3]					
WL.7.VAU						
ELA.7.39: [2]	12:(3)[2]	14:(1)[2]				
ELA.7.40						
ELA.7.41						

Table 7.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 CONF WV Grade 7B3 ELA (Three Reviewers)

Low		Medium		High	
1.8		5.4		9	
1	10199-1406	Exact:3			
2	10199-1407	Exact:6			
3	10199-1409	Exact:6			
4	10199-5226	Exact:2		Partial:1	
5	10199-5230	Exact:2		Partial:1	
6	10199-5231	Partial:2		Minimal:1	
7	10199-5233	Exact:2		Partial:1	
8	10199-5234	Exact:2		Partial:1	
9	10199-5235	Exact:1		Partial:1	
10	10199-5236	Exact:3		Minimal:1	
11	10199-5238	Exact:2		Minimal:1	
12	10199-5625	Exact:3			
13	10199-5831	Exact:3			
14	REP10199-5400	Exact:2		Partial:1	
15	REP10199-5401	Exact:3			
16	10199-5403	Exact:3			
17	REP10199-5405	Exact:3			
18	10199-5407	Exact:2		Minimal:1	
19	REP10199-5408	Exact:3			
20	REP10199-5409	Exact:3			
21	REP10199-6589	Exact:3			
22	REP10199-6604	Exact:2		Partial:1	
23	10199-7515	Exact:3			
24	REP10199-7518	Exact:2		Partial:1	
25	REP10199-7516	Exact:2		Partial:1	
26	REP10199-7510	Exact:1		Partial:2	
27	REP10199-7511	Exact:2		Partial:1	
28	REP10199-7514	Exact:3			
29	REP10199-7512	Exact:3			
30	REP10199-3240	Exact:3			

31 REP10199-3241	Exact:6		
32 REP10199-3242	Exact:3		
33 REP10199-3243	Exact:3		
34 10199-3910	Exact:3		
35 10199-3913	Exact:1	Partial:1	Minimal:1
36 10199-3911	Exact:1	Partial:2	
37 10199-4834	Exact:3		
38 10199-4837	Exact:3		
39 CONVENTIONS10199-14168	Minimal:6		
40 ELABORATION10199-14168	Exact:9		
41 ORGANIZATION10199-14168	Partial:9		

Grade 8 Batch 1 ELA West Virginia

Table 8.1 *Categorical Concurrence between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 8 B1 Number of Assessment Items - 42*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.8 Grade 8 Literary Reading	3	8.2	2 3	1 7	12.5 87.5	14.2	0.84	YES
IR.8 Grade 8 Informational Reading	3	9.8	2 3	2 7	22.22 77.78	16	0.71	YES
WL.8 Grade 8 Writing and Language	7	16	1 2 3 4	1 6 8 1	6.25 37.5 50 6.25	32.4	1.52	YES
Total	13	34	1 2 3 4	1 9 22 1	3 27 67 3	62.6	0.89	

Table 8.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 8 B1 Number of Assessment Items - 42*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.8 Grade 8 Literary Reading	3	8.2	14.2	0.84	71.23	19	27.44	19	1.33	3	NO
IR.8 Grade 8 Informational Reading	3	9.8	16	0.71	72.38	12	27.62	12	0	0	NO
WL.8 Grade 8 Writing and Language	7	16	32.4	1.52	16.51	15	70.18	9	13.32	9	YES
Total	13	34	62.6	0.89	43.13	13.8	49.52	10	7.35	4.7	

Table 8.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Five Reviewers
WV ELA 2019 Grade 8 B1 Number of Assessment Items - 42

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.8 Grade 8 Literary Reading	3	8.2	14.2	0.84	6.2	0.45	75.56	1.24	YES	32	2	0.71	0.05	YES
IR.8 Grade 8 Informational Reading	3	9.8	16	0.71	9	0.71	91.78	4.62	YES	36	2	0.81	0.03	YES
WL.8 Grade 8 Writing and Language	7	16	32.4	1.52	6	0	37.5	0	NO	33	3	0.74	0.05	YES
Total	13	34	62.6	0.89	7.1	1.68	68.28	28		34	2	0.75	0.05	

Table 8.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Five Reviewers
WV ELA 2019 Grade 8 B1 Number of Assessment Items - 42

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.8 Grade 8 Literary Reading	YES	NO	YES	YES
IR.8 Grade 8 Informational Reading	YES	NO	YES	YES
WL.8 Grade 8 Writing and Language	YES	YES	NO	YES

Table 8.5
Depth-of-Knowledge Levels by Item and Reviewers
Intraclass Correlation
WV ELA 2019 Grade 8 B1
Reviewer's DOK

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1	2	1	2	1	2
2	2	1	2	1	2
3	1	1	2	1	2
4	2	2	2	2	2
5	2	2	2	2	2
6	2	2	2	2	2
7	2	2	3	2	3
8	2	2	2	2	2
9	2	2	2	2	2
10	2	2	2	2	2
11	2	2	2	2	2
12	2	2	2	2	2
13	2	2	2	2	2
14	2	2	2	2	2
15	2	2	2	2	3
16	2	2	2	2	2
17	2	2	2	2	2
18	2	2	2	2	2
19	2	2	2	2	2
20	2	2	2	2	2
21	2	2	2	2	2
22	2	2	2	2	2
23	2	2	2	2	2
24	2	2	2	2	2
25	2	2	2	2	2
26	2	2	2	2	2
27	2	2	2	2	2
28	2	3	2	2	2
29	2	2	2	2	3

30	2	2	2	2	2
31	2	2	2	2	3
32	2	1	2	1	2
33	2	1	2	1	2
34	1	1	2	1	1
35	2	2	2	2	2
36	2	3	2	2	2
37	2	3	2	2	2
38	2	2	2	2	2
39	2	2	2	2	3
40	1	1	1	1	1
41	3	3	3	3	3
42	3	3	3	3	3

Intraclass correlation - .9039

Pairwise Comparison - 0.84

Table 8.6

DOK Levels and Objectives Code by Each Reviewer

WV ELA 2019 Grade 8 B1

Number of Reviewers: Five

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.8.38			1	ELA.8.38			2	ELA.8.38			1	ELA.8.36			2	ELA.8.38		
2	2	ELA.8.38			1	ELA.8.38			2	ELA.8.38			1	ELA.8.36			2	ELA.8.38		
3	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.36			2	ELA.8.37		
4	2	ELA.8.11			2	ELA.8.11			2	ELA.8.11			2	ELA.8.11			2	ELA.8.11		
5	2	ELA.8.39			2	ELA.8.39			2	ELA.8.39			2	ELA.8.39			2	ELA.8.39		
6	2	ELA.8.10			2	ELA.8.10			2	ELA.8.10			2	ELA.8.10			2	ELA.8.10		
7	2	ELA.8.16			2	ELA.8.16			3	ELA.8.16			2	ELA.8.16			3	ELA.8.16		
8	2	ELA.8.4			2	ELA.8.4			2	ELA.8.4			2	ELA.8.11			2	ELA.8.4		
9	2	ELA.8.4			2	ELA.8.4			2	ELA.8.4			2	ELA.8.4			2	ELA.8.4		
10	2	ELA.8.39			2	ELA.8.10			2	ELA.8.39			2	ELA.8.39			2	ELA.8.39		
11	2	ELA.8.6			2	ELA.8.6			2	ELA.8.6			2	ELA.8.17			2	ELA.8.6		
12	2	ELA.8.3			2	ELA.8.1			2	ELA.8.3			2	ELA.8.1			2	ELA.8.1		
13	2	ELA.8.7			2	ELA.8.7			2	ELA.8.39			2	ELA.8.7			2	ELA.8.39		
14	2	ELA.8.3			2	ELA.8.1			2	ELA.8.1			2	ELA.8.1			2	ELA.8.1		
15	2	ELA.8.14			2	ELA.8.14			2	ELA.8.14			2	ELA.8.8			3	ELA.8.14		
16	2	ELA.8.3			2	ELA.8.3			2	ELA.8.3			2	ELA.8.3			2	ELA.8.3		
17	2	ELA.8.7			2	ELA.8.7			2	ELA.8.1			2	ELA.8.3			2	ELA.8.7		
18	2	ELA.8.7			2	ELA.8.7			2	ELA.8.7			2	ELA.8.7			2	ELA.8.7		
19	2	ELA.8.2			2	ELA.8.2			2	ELA.8.1			2	ELA.8.2			2	ELA.8.1		
20	2	ELA.8.1			2	ELA.8.1			2	ELA.8.1			2	ELA.8.3			2	ELA.8.1		
21	2	ELA.8.7			2	ELA.8.7			2	ELA.8.39			2	ELA.8.39			2	ELA.8.39		
22	2	ELA.8.4			2	ELA.8.4			2	ELA.8.4			2	ELA.8.6			2	ELA.8.4		
23	2	ELA.8.5			2	ELA.8.5			2	ELA.8.4			2	ELA.8.5			2	ELA.8.5		
24	2	IR.8.CS			2	IR.8.CS			2	LR.8.CS			2	IR.8.CS			2	IR.8.CS		
25	2	ELA.8.6			2	ELA.8.6			2	ELA.8.6			2	ELA.8.12			2	ELA.8.6		
26	2	ELA.8.11			2	ELA.8.11			2	ELA.8.11			2	ELA.8.11			2	ELA.8.11		
27	2	ELA.8.12			2	ELA.8.12			2	ELA.8.12			2	ELA.8.12			2	ELA.8.12		
28	2	ELA.8.4			3	ELA.8.12			2	ELA.8.12			2	ELA.8.12			2	ELA.8.12		
29	2	ELA.8.15			2	ELA.8.15			2	ELA.8.15			2	ELA.8.15			3	ELA.8.15		
30	2	ELA.8.15			2	ELA.8.16			2	ELA.8.16			2	ELA.8.6			2	ELA.8.4		
31	2	ELA.8.16			2	ELA.8.16			2	ELA.8.5			2	ELA.8.16			3	ELA.8.16		

32	2	ELA.8.38		1	ELA.8.38		2	ELA.8.38		1	ELA.8.36		2	ELA.8.38		
33	2	ELA.8.38		1	ELA.8.38		2	ELA.8.38		1	ELA.8.36		2	ELA.8.38		
34	1	ELA.8.37		1	ELA.8.37		2	ELA.8.37		1	ELA.8.37		1	ELA.8.37		
35	2	ELA.8.3		2	ELA.8.3		2	ELA.8.3		2	ELA.8.3		2	ELA.8.3		
36	2	ELA.8.3		3	ELA.8.1		2	ELA.8.1		2	ELA.8.3		2	ELA.8.1		
37	2	ELA.8.2		3	ELA.8.2		2	ELA.8.2		2	ELA.8.2		2	ELA.8.2		
38	2	ELA.8.3		2	ELA.8.3		2	ELA.8.3		2	ELA.8.3		2	ELA.8.3		
39	2	ELA.8.9		2	ELA.8.9		2	ELA.8.9		2	ELA.8.9		3	ELA.8.9		
40	1	ELA.8.36	ELA.8.37	1	ELA.8.36	ELA.8.37	1	ELA.8.36	ELA.8.37	1	ELA.8.36	ELA.8.37	ELA.8.38	1	ELA.8.36	ELA.8.37
41	3	ELA.8.20		3	ELA.8.20		3	ELA.8.20		3	ELA.8.20	ELA.8.23		3	ELA.8.20	
42	3	ELA.8.23	ELA.8.38	ELA.8.20	3	ELA.8.20	ELA.8.23	ELA.8.38	3	ELA.8.20	ELA.8.23	ELA.8.38	3	ELA.8.20	ELA.8.23	ELA.8.38
Objective Pairwise Comparison: 0.69																
Standard Pairwise Comparison: 0.95																

Table 8.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV ELA 2019 Grade 8 B1

	Low		Medium				High	
	0		12				20	
LR.8								
LR.8.KID								
ELA.8.1	12(3)	14(4)	17(1)	20(4)	19(2)	36(3)		
ELA.8.2	37(5)	19(3)						
ELA.8.3	20(1)	17(1)	14(1)	16(5)	12(2)	38(5)	36(2)	35(5)
LR.8.CS	24(1)							
ELA.8.7	17(3)	18(5)	21(2)	13(3)				
ELA.8.8	15(1)							
ELA.8.9	39(5)							
LR.8.IKI								
ELA.8.13								
ELA.8.14	15(4)							
IR.8								
IR.8.KID								
ELA.8.4	9(5)	8(4)	23(1)	22(4)	28(1)	30(1)		
ELA.8.5	31(1)	23(4)						
ELA.8.6	22(1)	30(1)	25(4)	11(4)				
IR.8.CS	24(4)							
ELA.8.10	10(1)	6(5)						
ELA.8.11	8(1)	4(5)	26(5)					
ELA.8.12	27(5)	28(4)	25(1)					
IR.8.IKI								

ELA.8.15	30(1)	29(5)					
ELA.8.16	30(2)	31(4)	7(5)				
ELA.8.17	11(1)						
WL.8							
WL.8.TTP							
ELA.8.20	41(20)	42(20)					
ELA.8.21							
ELA.8.22							
WL.8.PDW							
ELA.8.23	41(4)	42(16)					
ELA.8.24							
ELA.8.25							
WL.8.RBPK							
ELA.8.26							
ELA.8.27							
ELA.8.28							
WL.8.RW							
ELA.8.29							
WL.8.CSE							
ELA.8.36	40(10)	33(2)	3(2)	1(1)	2(2)	32(1)	
ELA.8.37	3(8)	34(10)	40(10)				
WL.8.KL							
ELA.8.38	33(8)	2(8)	1(4)	32(4)	42(16)	40(2)	
WL.8.VAU							
ELA.8.39	21(3)	5(5)	10(4)	13(2)			
ELA.8.40							
ELA.8.41							

Table 8.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 8 B1

Low		Medium		High	
4		12		20	
1	10199-3253	ELA.8.36:1	ELA.8.38:4		
2	10199-3254	ELA.8.36:2	ELA.8.38:8		
3	10199-3255	ELA.8.36:2	ELA.8.37:8		
4	10199-3472	ELA.8.11:5			
5	10199-3476	ELA.8.39:5			
6	10199-3477	ELA.8.10:5			
7	10199-3478	ELA.8.16:5			
8	10199-3484	ELA.8.4:4	ELA.8.11:1		
9	10199-3593	ELA.8.4:5			
10	10199-3728	ELA.8.10:1	ELA.8.39:4		
11	10199-4822	ELA.8.6:4	ELA.8.17:1		
12	10199-5771	ELA.8.1:3	ELA.8.3:2		
13	REP10199-5782	ELA.8.7:3	ELA.8.39:2		
14	REP10199-5785	ELA.8.1:4	ELA.8.3:1		
15	REP10199-5788	ELA.8.8:1	ELA.8.14:4		
16	10199-5789	ELA.8.3:5			
17	REP10199-5791	ELA.8.1:1	ELA.8.3:1	ELA.8.7:3	
18	REP10199-5800	ELA.8.7:5			
19	REP10199-5805	ELA.8.1:2	ELA.8.2:3		
20	REP10199-6501	ELA.8.1:4	ELA.8.3:1		
21	REP10199-6502	ELA.8.7:2	ELA.8.39:3		
22	REP10199-7201	ELA.8.4:4	ELA.8.6:1		
23	REP10199-7203	ELA.8.4:1	ELA.8.5:4		
24	10199-7204	LR.8.CS:1	IR.8.CS:4		
25	REP10199-7205	ELA.8.6:4	ELA.8.12:1		
26	REP10199-7208	ELA.8.11:5			
27	REP10199-7209	ELA.8.12:5			
28	REP10199-7210	ELA.8.4:1	ELA.8.12:4		
29	REP10199-7211	ELA.8.15:5			
30	10199-7212	ELA.8.4:1	ELA.8.6:1	ELA.8.15:1	ELA.8.16:2

31 10199-7217	ELA.8.5:1	ELA.8.16:4	
32 10199-3159	ELA.8.36:1	ELA.8.38:4	
33 10199-3160	ELA.8.36:2	ELA.8.38:8	
34 10199-3161	ELA.8.37:10		
35 10199-7223	ELA.8.3:5		
36 REP10199-7220	ELA.8.1:3	ELA.8.3:2	
37 REP10199-7221	ELA.8.2:5		
38 10199-7222	ELA.8.3:5		
39 10199-7224	ELA.8.9:5		
40 CONVENTIONS10199-14170	ELA.8.36:10	ELA.8.37:10	ELA.8.38:2
41 ELABORATION10199-14170	ELA.8.20:20	ELA.8.23:4	
42 ORGANIZATION10199-14170	ELA.8.20:20	ELA.8.23:16	ELA.8.38:16

Table 8.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 8 B1

Low DOK		Matched DOK		High DOK

LR.8								
LR.8.KID								
ELA.8.1: [2]	12:(3)[2]	14:(4)[2]	17:(1)[2]	19:(2)[2]	20:(4)[2]	36:(3)[2]		
ELA.8.2: [3]	19:(3)[2]	37:(5)[2]						
ELA.8.3: [3]	12:(2)[2]	14:(1)[2]	16:(5)[2]	17:(1)[2]	20:(1)[2]	35:(5)[2]	36:(2)[2]	38:(5)[2]
LR.8.CS: [3]	24:(1)[2]							
ELA.8.7: [3]	13:(3)[2]	17:(3)[2]	18:(5)[2]	21:(2)[2]				
ELA.8.8: [3]	15:(1)[2]							
ELA.8.9: [3]	39:(5)[2]							
LR.8.IKI								
ELA.8.13								
ELA.8.14: [3]	15:(4)[2]							
IR.8								
IR.8.KID								
ELA.8.4: [2]	8:(4)[2]	9:(5)[2]	22:(4)[2]	23:(1)[2]	28:(1)[2]	30:(1)[2]		
ELA.8.5: [3]	23:(4)[2]	31:(1)[2]						
ELA.8.6: [3]	11:(4)[2]	22:(1)[2]	25:(4)[2]	30:(1)[2]				
IR.8.CS: [3]	24:(4)[2]							
ELA.8.10: [3]	6:(5)[2]	10:(1)[2]						
ELA.8.11: [3]	4:(5)[2]	8:(1)[2]	26:(5)[2]					
ELA.8.12: [3]	25:(1)[2]	27:(5)[2]	28:(4)[2]					
IR.8.IKI								
ELA.8.15: [3]	29:(5)[2]	30:(1)[2]						
ELA.8.16: [3]	7:(5)[2]	30:(2)[2]	31:(4)[2]					
ELA.8.17: [2]	11:(1)[2]							
WL.8								
WL.8.TTP								
ELA.8.20: [3]	41:(20)[3]	42:(20)[3]						

ELA.8.21								
ELA.8.22								
WL.8.PDW								
ELA.8.23: [3]	41:(4)[3]	42:(16)[3]						
ELA.8.24								
ELA.8.25								
WL.8.RBPK								
ELA.8.26								
ELA.8.27								
ELA.8.28								
WL.8.RW								
ELA.8.29								
WL.8.CSE								
ELA.8.36: [2]	1:(1)[1]	2:(2)[1]	3:(2)[1]	32:(1)[1]	33:(2)[1]	40:(10)[1]		
ELA.8.37: [1]	3:(8)[2]	34:(10)[1]	40:(10)[1]					
WL.8.KL								
ELA.8.38: [2]	1:(4)[2]	2:(8)[2]	32:(4)[2]	33:(8)[2]	40:(2)[1]	42:(16)[3]		
WL.8.VAU								
ELA.8.39: [2]	5:(5)[2]	10:(4)[2]	13:(2)[2]	21:(3)[2]				
ELA.8.40								
ELA.8.41								

Table 8.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 CONF WV Grade 8B1 ELA

Low		Medium		High	
4		12		20	
1	10199-3253	Exact:6			
2	10199-3254	Exact:12			
3	10199-3255	Exact:12			
4	10199-3472	Exact:4		Partial:2	
5	10199-3476	Exact:4		Partial:2	
6	10199-3477	Exact:5		Partial:1	
7	10199-3478	Exact:4		Partial:2	
8	10199-3484	Exact:5		Partial:1	
9	10199-3593	Exact:4		Partial:2	
10	10199-3728	Exact:3		Partial:3	
11	10199-4822	Exact:4		Partial:2	
12	10199-5771	Exact:4		Partial:2	
13	REP10199-5782	Exact:2		Partial:3	
14	REP10199-5785	Exact:6			
15	REP10199-5788	Exact:5		Partial:1	
16	10199-5789	Exact:6			
17	REP10199-5791	Exact:3		Partial:2	
18	REP10199-5800	Exact:3		Partial:3	
19	REP10199-5805	Exact:4		Partial:2	
20	REP10199-6501	Exact:5		Partial:1	
21	REP10199-6502	Exact:3		Partial:3	
22	REP10199-7201	Exact:5		Partial:1	
23	REP10199-7203	Exact:4		Partial:2	
24	10199-7204	Partial:4		Minimal:1	
25	REP10199-7205	Exact:2		Partial:4	
26	REP10199-7208	Exact:4		Partial:2	
27	REP10199-7209	Exact:4		Partial:2	
28	REP10199-7210	Exact:6			
29	REP10199-7211	Exact:6			
30	10199-7212	Exact:3		Partial:3	

31 10199-7217	Partial:3	Negligible:2	
32 10199-3159	Exact:5		
33 10199-3160	Exact:10		
34 10199-3161	Exact:10		
35 10199-7223	Exact:5		
36 REP10199-7220	Exact:3	Partial:2	
37 REP10199-7221	Exact:5		
38 10199-7222	Exact:5		
39 10199-7224	Exact:5		
40 CONVENTIONS10199-14170	Minimal:10		
41 ELABORATION10199-14170	Exact:20		
42 ORGANIZATION10199-14170	Partial:20		

Grade 8 Batch 2 ELA West Virginia

Table 8.1 *Categorical Concurrence between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 8 B2 Number of Assessment Items - 42*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.8 Grade 8 Literary Reading	3	8	2 3	1 7	12.5 87.5	12.4	2.07	YES
IR.8 Grade 8 Informational Reading	3	9	2 3	2 7	22.22 77.78	18.6	1.52	YES
WL.8 Grade 8 Writing and Language	7	16	1 2 3 4	1 6 8 1	6.25 37.5 50 6.25	30.2	1.79	YES
Total	13	33	1 2 3 4	1 9 22 1	3 27 67 3	61.2	1.79	

Table 8.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 8 B2 Number of Assessment Items - 42*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.8 Grade 8 Literary Reading	3	8	12.4	2.07	69.29	14	30.71	14	0	0	NO
IR.8 Grade 8 Informational Reading	3	9	18.6	1.52	69.8	13	28.04	12	2.16	3	NO
WL.8 Grade 8 Writing and Language	7	16	30.2	1.79	13.5	7	71.11	6	15.39	11	YES
Total	13	33	61.2	1.79	41.83	7.5	49.67	3.7	8.5	6	

*Table 8.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Five Reviewers
WV ELA 2019 Grade 8 B2 Number of Assessment Items - 42*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.8 Grade 8 Literary Reading	3	8	12.4	2.07	5.8	0.45	72.5	5.59	YES	28	4	0.76	0.04	YES
IR.8 Grade 8 Informational Reading	3	9	18.6	1.52	7.6	0.89	84.44	9.94	YES	39	4	0.8	0.05	YES
WL.8 Grade 8 Writing and Language	7	16	30.2	1.79	5.4	0.55	33.75	3.42	NO	33	3	0.81	0.02	YES
Total	13	33	61.2	1.79	6.3	1.17	63.56	27		33	6	0.79	0.03	

*Table 8.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Five Reviewers WV ELA 2019 Grade 8 B2
Number of Assessment Items - 42*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.8 Grade 8 Literary Reading	YES	NO	YES	YES
IR.8 Grade 8 Informational Reading	YES	NO	YES	YES
WL.8 Grade 8 Writing and Language	YES	YES	NO	YES

Table 8.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 8 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1	2	1	2	1	2
2	2	1	2	1	2
3	1	1	2	1	1
4	1	1	2	1	2
5	2	2	2	2	2
6	2	2	2	2	2
7	2	3	2	2	2
8	2	2	2	2	2
9	2	2	2	2	2
10	2	2	2	2	2
11	2	2	2	2	3
12	3	2	2	2	2
13	2	2	2	2	3
14	3	3	2	2	2
15	2	2	2	2	2
16	2	2	2	2	2
17	2	2	2	2	3
18	2	2	2	2	2
19	2	2	2	2	2
20	2	1	2	2	2
21	2	2	2	2	2
22	2	2	2	2	2
23	2	2	2	2	2
24	2	2	2	2	2
25	2	2	2	2	2
26	2	3	2	2	3
27	2	2	2	2	2
28	2	2	1	2	3
29	2	3	2	2	2
30	2	2	2	2	2
31	2	2	2	2	2

32	1	1	2	1	1
33	1	1	2	1	1
34	1	1	2	1	1
35	1	1	2	1	1
36	2	2	2	2	2
37	2	2	2	2	2
38	2	2	2	2	2
39	2	3	3	2	2
40	1	1	1	1	2
41	3	3	3	3	3
42	3	3	3	3	3

Intraclass correlation - .8978

Pairwise Comparison - 0.77

Table 8.6

*DOK Levels and Objectives Code by Each Reviewer**WV ELA 2019 Grade 8 B2*

Number of Reviewers: Five

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.8.38			1	ELA.8.38			2	ELA.8.38			1	ELA.8.38			2	ELA.8.38		
2	2	ELA.8.38			1	ELA.8.38			2	ELA.8.38			1	ELA.8.38			2	ELA.8.38		
3	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			1	ELA.8.37		
4	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			2	ELA.8.4		
5	2	ELA.8.4			2	ELA.8.4			2	ELA.8.4			2	ELA.8.17			2	ELA.8.5		
6	2	ELA.8.5			2	ELA.8.5			2	ELA.8.5			2	ELA.8.5			2	ELA.8.5		
7	2	ELA.8.5			3	ELA.8.5			2	ELA.8.5			2	ELA.8.5			2	ELA.8.6		
8	2	ELA.8.6			2	ELA.8.6			2	ELA.8.6			2	ELA.8.12			2	ELA.8.10		
9	2	ELA.8.10			2	ELA.8.10			2	ELA.8.10			2	ELA.8.10			2	ELA.8.11		
10	2	ELA.8.11			2	ELA.8.11			2	ELA.8.11			2	ELA.8.11			2	ELA.8.12		
11	2	ELA.8.12			2	ELA.8.12			2	ELA.8.12			2	ELA.8.12			3	ELA.8.17		
12	3	ELA.8.17			2	ELA.8.17			2	ELA.8.17			2	ELA.8.12			2	ELA.8.10		
13	2	ELA.8.10			2	ELA.8.10			2	ELA.8.39			2	ELA.8.10			3	ELA.8.11		
14	3	ELA.8.11			3	ELA.8.11			2	ELA.8.11			2	ELA.8.11			2	ELA.8.3		
15	2	ELA.8.3			2	ELA.8.3			2	ELA.8.3			2	ELA.8.3			2	ELA.8.39		
16	2	ELA.8.7			2	ELA.8.7			2	ELA.8.39			2	ELA.8.39			2	ELA.8.1		
17	2	ELA.8.3			2	ELA.8.1			2	ELA.8.1			2	ELA.8.1			3	ELA.8.14		
18	2	ELA.8.14			2	ELA.8.14			2	ELA.8.14			2	ELA.8.14			2	ELA.8.7		
19	2	ELA.8.7			2	ELA.8.7			2	ELA.8.1			2	ELA.8.3			2	ELA.8.7		
20	2	ELA.8.7			1	ELA.8.7			2	ELA.8.7			2	ELA.8.7			2	ELA.8.7		
21	2	ELA.8.3			2	ELA.8.3			2	ELA.8.3			2	ELA.8.3			2	ELA.8.3		
22	2	ELA.8.2			2	ELA.8.2			2	ELA.8.1			2	ELA.8.2			2	ELA.8.1		
23	2	ELA.8.1			2	ELA.8.1			2	ELA.8.1			2	ELA.8.3			2	ELA.8.1		
24	2	ELA.8.7			2	ELA.8.7			2	ELA.8.39			2	ELA.8.39			2	ELA.8.39		
25	2	ELA.8.4			2	ELA.8.4			2	ELA.8.4			2	ELA.8.4			2	ELA.8.4		
26	2	ELA.8.11			3	ELA.8.11			2	ELA.8.11			2	ELA.8.11			3	ELA.8.11		
27	2	ELA.8.5			2	ELA.8.12			2	ELA.8.12			2	ELA.8.12			2	ELA.8.12		
28	2	ELA.8.15			2	ELA.8.15			1	ELA.8.15			2	ELA.8.17			3	ELA.8.15		
29	2	ELA.8.5			3	ELA.8.5			2	ELA.8.5			2	ELA.8.5			2	ELA.8.5		
30	2	ELA.8.10			2	ELA.8.10			2	ELA.8.10			2	ELA.8.11			2	ELA.8.10		

31	2	ELA.8.4			2	ELA.8.6			2	ELA.8.6			2	ELA.8.4			2	ELA.8.6		
32	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			1	ELA.8.37		
33	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			1	ELA.8.37		
34	1	ELA.8.38			1	ELA.8.38			2	ELA.8.38			1	ELA.8.38			1	ELA.8.39		
35	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			1	ELA.8.37		
36	2	ELA.8.9			2	ELA.8.7			2	ELA.8.7			2	ELA.8.1			2	ELA.8.7		
37	2	ELA.8.1			2	ELA.8.1			2	ELA.8.1			2	ELA.8.11			2	ELA.8.1		
38	2	ELA.8.2			2	ELA.8.2			2	ELA.8.2			2	ELA.8.5			2	ELA.8.2		
39	2	ELA.8.9			3	ELA.8.9			3	ELA.8.9			2	ELA.8.12			2	ELA.8.9		
40	1	ELA.8.36	ELA.8.37		1	ELA.8.36	ELA.8.37		1	ELA.8.36	ELA.8.37		1	ELA.8.38	ELA.8.37		2	ELA.8.36	ELA.8.37	
41	3	ELA.8.21			3	ELA.8.21			3	ELA.8.21			3	ELA.8.21	ELA.8.23		3	ELA.8.21		
42	3	ELA.8.21	ELA.8.23	ELA.8.38	3	ELA.8.21	ELA.8.23	ELA.8.38	3	ELA.8.21	ELA.8.23	ELA.8.38	3	ELA.8.21			3	ELA.8.21	ELA.8.23	ELA.8.36
Objective Pairwise Comparison: 0.63																				
Standard Pairwise Comparison: 0.9																				

Table 8.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV ELA 2019 Grade 8 B2

	Low		Medium		High		
	0		12			20	
LR.8							
LR.8.KID							
ELA.8.1	16(1)	17(3)	19(1)	22(2)	23(4)	36(1)	37(4)
ELA.8.2	38(4)	22(3)					
ELA.8.3	21(5)	19(1)	23(1)	17(1)	15(4)	14(1)	
LR.8.CS							
ELA.8.7	16(2)	18(1)	24(2)	19(3)	20(5)	36(3)	
ELA.8.8							
ELA.8.9	36(1)	39(4)					
LR.8.IKI							
ELA.8.13							
ELA.8.14	18(4)	17(1)					
IR.8							
IR.8.KID							
ELA.8.4	5(3)	4(2)	25(5)	31(2)			
ELA.8.5	27(1)	29(5)	6(5)	7(4)	5(1)	38(1)	
ELA.8.6	7(1)	8(3)	31(3)				
IR.8.CS							
ELA.8.10	30(4)	8(1)	9(4)	12(1)	13(3)		
ELA.8.11	13(1)	14(4)	9(1)	10(4)	30(1)	26(10)	37(1)
ELA.8.12	39(1)	27(4)	10(1)	11(4)	8(1)	12(1)	
IR.8.IKI							

ELA.8.15	28(4)						
ELA.8.16							
ELA.8.17	28(1)	12(3)	5(1)	11(1)			
WL.8							
WL.8.TTP							
ELA.8.20							
ELA.8.21	41(20)	42(20)					
ELA.8.22							
WL.8.PDW							
ELA.8.23	41(4)	42(16)					
ELA.8.24							
ELA.8.25							
WL.8.RBPK							
ELA.8.26							
ELA.8.27							
ELA.8.28							
WL.8.RW							
ELA.8.29							
WL.8.CSE							
ELA.8.36	40(8)	42(4)					
ELA.8.37	35(5)	4(8)	3(5)	32(5)	33(5)	40(10)	
WL.8.KL							
ELA.8.38	34(8)	1(5)	2(5)	40(2)	42(12)		
WL.8.VAU							
ELA.8.39	13(1)	16(2)	15(1)	34(2)	24(3)		
ELA.8.40							
ELA.8.41							

Table 8.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 8 B2

Low		Medium		High	
4		12		20	
1 REP10199-2335	ELA.8.38:5				
2 REP10199-2336	ELA.8.38:5				
3 REP10199-2339	ELA.8.37:5				
4 REP10199-2340	ELA.8.4:2	ELA.8.37:8			
5 10199-5835	ELA.8.4:3	ELA.8.5:1	ELA.8.17:1		
6 10199-5837	ELA.8.5:5				
7 10199-5838	ELA.8.5:4	ELA.8.6:1			
8 10199-5839	ELA.8.6:3	ELA.8.10:1	ELA.8.12:1		
9 10199-5842	ELA.8.10:4	ELA.8.11:1			
10 10199-5844	ELA.8.11:4	ELA.8.12:1			
11 10199-5847	ELA.8.12:4	ELA.8.17:1			
12 10199-5851	ELA.8.10:1	ELA.8.12:1	ELA.8.17:3		
13 10199-5852	ELA.8.10:3	ELA.8.11:1	ELA.8.39:1		
14 10199-5855	ELA.8.3:1	ELA.8.11:4			
15 10199-5776	ELA.8.3:4	ELA.8.39:1			
16 REP10199-5782	ELA.8.1:1	ELA.8.7:2	ELA.8.39:2		
17 REP10199-5785	ELA.8.1:3	ELA.8.3:1	ELA.8.14:1		
18 REP10199-5788	ELA.8.7:1	ELA.8.14:4			
19 REP10199-5791	ELA.8.1:1	ELA.8.3:1	ELA.8.7:3		
20 REP10199-5800	ELA.8.7:5				
21 10199-5803	ELA.8.3:5				
22 REP10199-5805	ELA.8.1:2	ELA.8.2:3			
23 REP10199-6501	ELA.8.1:4	ELA.8.3:1			
24 REP10199-6502	ELA.8.7:2	ELA.8.39:3			
25 REP10199-8235	ELA.8.4:5				
26 REP10199-8291	ELA.8.11:10				
27 10199-8292	ELA.8.5:1	ELA.8.12:4			
28 10199-8293	ELA.8.15:4	ELA.8.17:1			
29 REP10199-8294	ELA.8.5:5				
30 REP10199-8295	ELA.8.10:4	ELA.8.11:1			

31 REP10199-13994	ELA.8.4:2	ELA.8.6:3	
32 REP10199-1262	ELA.8.37:5		
33 REP10199-1263	ELA.8.37:5		
34 REP10199-1264	ELA.8.38:8	ELA.8.39:2	
35 REP10199-1265	ELA.8.37:5		
36 10199-5528	ELA.8.1:1	ELA.8.7:3	ELA.8.9:1
37 10199-5526	ELA.8.1:4	ELA.8.11:1	
38 10199-5525	ELA.8.2:4	ELA.8.5:1	
39 10199-6492	ELA.8.9:4	ELA.8.12:1	
40 CONVENTIONS10199-14171	ELA.8.36:8	ELA.8.37:10	ELA.8.38:2
41 ELABORATION10199-14171	ELA.8.21:20	ELA.8.23:4	
42 ORGANIZATION10199-14171	ELA.8.21:20	ELA.8.23:16	ELA.8.36:4 ELA.8.38:12

Table 8.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 8 B2

Low DOK		Matched DOK		High DOK

LR.8							
LR.8.KID							
ELA.8.1: [2]	16:(1)[2]	17:(3)[2]	19:(1)[2]	22:(2)[2]	23:(4)[2]	36:(1)[2]	37:(4)[2]
ELA.8.2: [3]	22:(3)[2]	38:(4)[2]					
ELA.8.3: [3]	14:(1)[2]	15:(4)[2]	17:(1)[2]	19:(1)[2]	21:(5)[2]	23:(1)[2]	
LR.8.CS							
ELA.8.7: [3]	16:(2)[2]	18:(1)[2]	19:(3)[2]	20:(5)[2]	24:(2)[2]	36:(3)[2]	
ELA.8.8							
ELA.8.9: [3]	36:(1)[2]	39:(4)[2]					
LR.8.IKI							
ELA.8.13							
ELA.8.14: [3]	17:(1)[3]	18:(4)[2]					
IR.8							
IR.8.KID							
ELA.8.4: [2]	4:(2)[2]	5:(3)[2]	25:(5)[2]	31:(2)[2]			
ELA.8.5: [3]	5:(1)[2]	6:(5)[2]	7:(4)[2]	27:(1)[2]	29:(5)[2]	38:(1)[2]	
ELA.8.6: [3]	7:(1)[2]	8:(3)[2]	31:(3)[2]				
IR.8.CS							
ELA.8.10: [3]	8:(1)[2]	9:(4)[2]	12:(1)[2]	13:(3)[2]	30:(4)[2]		
ELA.8.11: [3]	9:(1)[2]	10:(4)[2]	13:(1)[3]	14:(4)[2]	26:(10)[2]	30:(1)[2]	37:(1)[2]
ELA.8.12: [3]	8:(1)[2]	10:(1)[2]	11:(4)[2]	12:(1)[2]	27:(4)[2]	39:(1)[2]	
IR.8.IKI							
ELA.8.15: [3]	28:(4)[2]						
ELA.8.16							
ELA.8.17: [2]	5:(1)[2]	11:(1)[3]	12:(3)[2]	28:(1)[2]			
WL.8							
WL.8.TTP							
ELA.8.20							

ELA.8.21: [3]	41:(20)[3]	42:(20)[3]					
ELA.8.22							
WL.8.PDW							
ELA.8.23: [3]	41:(4)[3]	42:(16)[3]					
ELA.8.24							
ELA.8.25							
WL.8.RBPK							
ELA.8.26							
ELA.8.27							
ELA.8.28							
WL.8.RW							
ELA.8.29							
WL.8.CSE							
ELA.8.36: [2]	40:(8)[1]	42:(4)[3]					
ELA.8.37: [1]	3:(5)[1]	4:(8)[1]	32:(5)[1]	33:(5)[1]	35:(5)[1]	40:(10)[1]	
WL.8.KL							
ELA.8.38: [2]	1:(5)[2]	2:(5)[2]	34:(8)[1]	40:(2)[1]	42:(12)[3]		
WL.8.VAU							
ELA.8.39: [2]	13:(1)[2]	15:(1)[2]	16:(2)[2]	24:(3)[2]	34:(2)[1]		
ELA.8.40							
ELA.8.41							

Table 8.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 8B2 ELA

Low		Medium		High	
4.8		14.4		24	
1 REP10199-2335		Exact:6			
2 REP10199-2336		Exact:6			
3 REP10199-2339		Exact:6			
4 REP10199-2340		Exact:12			
5 10199-5835		Exact:5		Minimal:1	
6 10199-5837		Exact:5		Partial:1	
7 10199-5838		Exact:5		Partial:1	
8 10199-5839		Exact:3		Partial:3	
9 10199-5842		Exact:4		Partial:2	
10 10199-5844		Exact:4		Partial:1	
11 10199-5847		Exact:4		Partial:1	
12 10199-5851		Exact:4		Partial:1	
13 10199-5852		Exact:2		Partial:3	
14 10199-5855		Exact:5		Partial:1	
15 10199-5776		Exact:4		Partial:2	
16 REP10199-5782		Exact:3		Partial:1	
17 REP10199-5785		Exact:5		Partial:1	
18 REP10199-5788		Exact:5		Partial:1	
19 REP10199-5791		Exact:4		Partial:1	
20 REP10199-5800		Exact:4		Partial:2	
21 10199-5803		Exact:4		Partial:2	
22 REP10199-5805		Exact:3		Partial:3	
23 REP10199-6501		Exact:5		Partial:1	
24 REP10199-6502		Exact:4		Partial:2	
25 REP10199-8235		Exact:5		Partial:1	
26 REP10199-8291		Exact:10		Partial:2	
27 10199-8292		Exact:6			
28 10199-8293		Exact:4		Partial:2	
29 REP10199-8294		Exact:6			
30 REP10199-8295		Exact:2		Partial:3	

31 REP10199-13994	Exact:1	Partial:5	
32 REP10199-1262	Exact:6		
33 REP10199-1263	Exact:6		
34 REP10199-1264	Exact:12		
35 REP10199-1265	Exact:6		
36 10199-5528	Exact:6		
37 10199-5526	Exact:4	Partial:2	
38 10199-5525	Exact:6		
39 10199-6492	Exact:6		
40 CONVENTIONS10199-14171	Exact:2	Minimal:10	
41 ELABORATION10199-14171	Exact:24		
42 ORGANIZATION10199-14171	Exact:4	Partial:20	

Grade 8 Batch 3 ELA West Virginia

Table 8.1 *Categorical Concurrence between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 8 B3 Number of Assessment Items - 43*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
LR.8 Grade 8 Literary Reading	3	8.4	2 3	1 7	12.5 87.5	14.8	0.84	YES
IR.8 Grade 8 Informational Reading	3	9.4	2 3	2 7	22.22 77.78	16	1.22	YES
WL.8 Grade 8 Writing and Language	7	16	1 2 3 4	1 6 8 1	6.25 37.5 50 6.25	32.2	1.64	YES
Total	13	33.8	1 2 3 4	1 9 22 1	3 27 67 3	63	0	

Table 8.2 *Depth-of-Knowledge Consistency Between Standards and Assessment as Rated by Five Reviewers WV ELA 2019 Grade 8 B3 Number of Assessment Items - 43*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
LR.8 Grade 8 Literary Reading	3	8.4	14.8	0.84	61.68	31	34.32	30	4	9	NO
IR.8 Grade 8 Informational Reading	3	9.4	16	1.22	69.41	14	29.16	12	1.43	3	NO
WL.8 Grade 8 Writing and Language	7	16	32.2	1.64	13.98	8	71.56	9	14.46	11	YES
Total	13	33.8	63	0	39.68	9.8	51.75	7.9	8.57	5.7	

*Table 8.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Five Reviewers
WV ELA 2019 Grade 8 B3 Number of Assessment Items - 43*

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
LR.8 Grade 8 Literary Reading	3	8.4	14.8	0.84	6.4	0.55	76.11	1.52	YES	32	2	0.82	0.08	YES
IR.8 Grade 8 Informational Reading	3	9.4	16	1.22	7.8	0.84	83.11	8.87	YES	33	3	0.76	0.03	YES
WL.8 Grade 8 Writing and Language	7	16	32.2	1.64	5.8	0.45	36.25	2.8	NO	35	4	0.79	0.05	YES
Total	13	33.8	63	0	6.7	1.03	65.16	25		33	2	0.79	0.03	

*Table 8.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Five Reviewers WV ELA 2019 Grade 8 B3
Number of Assessment Items - 43*

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
LR.8 Grade 8 Literary Reading	YES	NO	YES	YES
IR.8 Grade 8 Informational Reading	YES	NO	YES	YES
WL.8 Grade 8 Writing and Language	YES	YES	NO	YES

Table 8.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV ELA 2019 Grade 8 B3

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
1	2	1	2	2	2
2	2	1	2	2	2
3	2	1	2	2	2
4	2	4	2	2	2
5	2	4	2	2	2
6	2	4	2	2	2
7	2	4	2	2	2
8	2	3	2	2	2
9	2	2	2	2	3
10	2	2	2	2	3
11	2	2	2	2	3
12	2	2	2	2	2
13	2	2	2	2	3
14	2	2	2	2	2
15	2	2	2	2	3
16	2	2	2	2	2
17	2	2	2	2	3
18	2	2	2	2	3
19	2	2	2	2	3
20	3	3	2	2	3
21	1	1	2	1	1
22	1	1	2	1	1
23	2	1	2	1	1
24	1	1	2	1	1
25	2	2	2	2	2
26	2	2	2	2	2
27	2	3	2	2	3
28	2	3	2	2	2
29	2	2	2	2	2
30	2	2	2	2	2
31	2	2	2	2	3

32	2	2	2	2	2
33	2	2	2	2	2
34	2	2	2	2	3
35	2	3	2	2	3
36	2	2	3	2	2
37	2	1	2	2	2
38	2	1	2	1	2
39	1	1	2	1	2
40	1	1	2	1	2
41	1	1	1	1	1
42	3	3	3	3	3
43	3	3	3	3	3

Intraclass correlation - .8342

Pairwise Comparison - 0.68

Table 8.6

DOK Levels and Objectives Code by Each Reviewer

WV ELA 2019 Grade 8 B3

Number of Reviewers: Five

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	ELA.8.4			1	ELA.8.36			2	ELA.8.4			2	ELA.8.3			2	ELA.8.1		
2	2	ELA.8.5			1	ELA.8.36			2	ELA.8.4			2	ELA.8.5			2	ELA.8.5		
3	2	ELA.8.4			1	ELA.8.37			2	ELA.8.6			2	ELA.8.12			2	ELA.8.6		
4	2	ELA.8.7			4	ELA.8.2			2	ELA.8.10			2	ELA.8.10			2	ELA.8.10		
5	2	ELA.8.7			4	ELA.8.2			2	ELA.8.39			2	ELA.8.39			2	ELA.8.39		
6	2	ELA.8.11			4	ELA.8.2			2	ELA.8.11			2	ELA.8.11			2	ELA.8.11		
7	2	ELA.8.12			4	ELA.8.10			2	ELA.8.12			2	ELA.8.12			2	ELA.8.12		
8	2	ELA.8.12			3	ELA.8.12			2	ELA.8.12			2	ELA.8.12			2	ELA.8.12		
9	2	ELA.8.15			2	ELA.8.15			2	ELA.8.15			2	ELA.8.15			3	ELA.8.15		
10	2	ELA.8.16			2	ELA.8.16			2	ELA.8.16			2	ELA.8.16			3	ELA.8.16		
11	2	IR.8.CS			2	IR.8.CS			2	LR.8.CS			2	LR.8.CS			3	ELA.8.9		
12	2	ELA.8.1			2	ELA.8.1			2	ELA.8.1			2	ELA.8.1			2	ELA.8.1		
13	2	ELA.8.2			2	ELA.8.2			2	ELA.8.2			2	ELA.8.14			3	ELA.8.2		
14	2	ELA.8.3			2	ELA.8.3			2	ELA.8.3			2	ELA.8.3			2	ELA.8.3		
15	2	ELA.8.7			2	ELA.8.9			2	ELA.8.9			2	ELA.8.3			3	ELA.8.9		
16	2	ELA.8.7			2	ELA.8.7			2	ELA.8.7			2	ELA.8.7			2	ELA.8.39		
17	2	ELA.8.14			2	ELA.8.39			2	ELA.8.7			2	ELA.8.39			3	ELA.8.14		
18	2	ELA.8.3			2	ELA.8.14			2	ELA.8.14			2	ELA.8.3			3	ELA.8.14		
19	2	ELA.8.3			2	ELA.8.4			2	ELA.8.3			2	ELA.8.3			3	ELA.8.3		
20	3	ELA.8.14			3	ELA.8.14			2	ELA.8.14			2	ELA.8.14			3	ELA.8.14		
21	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			1	ELA.8.37		
22	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			1	ELA.8.37		
23	2	ELA.8.38			1	ELA.8.38			2	ELA.8.38			1	ELA.8.38			1	ELA.8.39		
24	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			1	ELA.8.37		
25	2	ELA.8.4			2	ELA.8.4			2	ELA.8.4			2	ELA.8.4			2	ELA.8.4		
26	2	ELA.8.5			2	ELA.8.11			2	ELA.8.11			2	ELA.8.11			2	ELA.8.5		
27	2	ELA.8.11			3	ELA.8.11			2	ELA.8.11			2	ELA.8.11			3	ELA.8.11		
28	2	ELA.8.5			3	ELA.8.5			2	ELA.8.5			2	ELA.8.5			2	ELA.8.5		
29	2	ELA.8.4			2	ELA.8.10			2	ELA.8.10			2	ELA.8.11			2	ELA.8.10		
30	2	ELA.8.4			2	ELA.8.6			2	ELA.8.6			2	ELA.8.4			2	ELA.8.6		

31	2	ELA.8.11			2	ELA.8.11			2	ELA.8.4			2	ELA.8.5			3	ELA.8.11		
32	2	ELA.8.1			2	ELA.8.1			2	ELA.8.1			2	ELA.8.3			2	ELA.8.1		
33	2	ELA.8.7			2	ELA.8.1			2	ELA.8.7			2	ELA.8.39			2	ELA.8.7		
34	2	ELA.8.7			2	ELA.8.7			2	ELA.8.1			2	ELA.8.7			3	ELA.8.7		
35	2	ELA.8.2			3	ELA.8.2			2	ELA.8.2			2	ELA.8.2			3	ELA.8.2		
36	2	ELA.8.9			2	ELA.8.9			3	ELA.8.9			2	ELA.8.9			2	ELA.8.38		
37	2	ELA.8.38			1	ELA.8.38			2	ELA.8.38			2	ELA.8.38			2	ELA.8.38		
38	2	ELA.8.38			1	ELA.8.38			2	ELA.8.38			1	ELA.8.38			2	ELA.8.38		
39	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			2	ELA.8.37		
40	1	ELA.8.37			1	ELA.8.37			2	ELA.8.37			1	ELA.8.37			2	ELA.8.37		
41	1	ELA.8.36	ELA.8.37		1	ELA.8.36	ELA.8.37		1	ELA.8.36	ELA.8.37		1	ELA.8.36	ELA.8.37		1	ELA.8.36	ELA.8.37	
42	3	ELA.8.20			3	ELA.8.20			3	ELA.8.20			3	ELA.8.20	ELA.8.23		3	ELA.8.20		
43	3	ELA.8.20	ELA.8.23	ELA.8.38	3	ELA.8.20	ELA.8.23	ELA.8.38	3	ELA.8.20	ELA.8.23	ELA.8.38	3	ELA.8.20	ELA.8.23		3	ELA.8.20	ELA.8.23	ELA.8.36
Objective Pairwise Comparison: 0.65																				
Standard Pairwise Comparison: 0.86																				

Table 8.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV ELA 2019 Grade 8 B3

Low	Medium				High			
0	12				20			
LR.8								
LR.8.KID								
ELA.8.1	1(1)	12(5)	32(4)	33(1)	34(1)			
ELA.8.2	35(5)	13(4)	5(1)	6(1)	4(1)			
ELA.8.3	1(1)	14(5)	15(1)	32(1)	18(2)	19(4)		
LR.8.CS	11(2)							
ELA.8.7	16(4)	15(1)	5(1)	4(1)	17(1)	33(3)	34(4)	
ELA.8.8								
ELA.8.9	36(4)	15(3)	11(1)					
LR.8.IKI								
ELA.8.13								
ELA.8.14	13(1)	17(2)	18(3)	20(5)				
IR.8								
IR.8.KID								
ELA.8.4	19(1)	30(2)	31(1)	25(5)	29(1)	3(1)	1(2)	2(1)
ELA.8.5	2(3)	28(5)	26(2)	31(1)				
ELA.8.6	30(3)	3(2)						
IR.8.CS	11(2)							
ELA.8.10	4(3)	7(1)	29(3)					
ELA.8.11	29(1)	26(3)	27(10)	31(3)	6(4)			
ELA.8.12	7(4)	8(5)	3(1)					
IR.8.IKI								
ELA.8.15	9(5)							
ELA.8.16	10(5)							
ELA.8.17								

WL.8								
WL.8.TTP								
ELA.8.20	42(20)	43(20)						
ELA.8.21								
ELA.8.22								
WL.8.PDW								
ELA.8.23	42(4)	43(20)						
ELA.8.24								
ELA.8.25								
WL.8.RBPK								
ELA.8.26								
ELA.8.27								
ELA.8.28								
WL.8.RW								
ELA.8.29								
WL.8.CSE								
ELA.8.36	41(10)	2(1)	1(1)	43(4)				
ELA.8.37	3(1)	24(5)	21(5)	22(5)	39(5)	40(5)	41(10)	
WL.8.KL								
ELA.8.38	36(1)	37(5)	38(10)	23(8)	43(12)			
WL.8.VAU								
ELA.8.39	23(2)	17(2)	33(1)	5(3)	16(1)			
ELA.8.40								
ELA.8.41								

Table 8.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV ELA 2019 Grade 8 B3

Low		Medium		High	
4		12		20	
1 REP10199-7201	ELA.8.1:1	ELA.8.3:1	ELA.8.4:2	ELA.8.36:1	
2 REP10199-7203	ELA.8.4:1	ELA.8.5:3	ELA.8.36:1		
3 REP10199-7205	ELA.8.4:1	ELA.8.6:2	ELA.8.12:1	ELA.8.37:1	
4 10199-7206	ELA.8.2:1	ELA.8.7:1	ELA.8.10:3		
5 10199-7207	ELA.8.2:1	ELA.8.7:1	ELA.8.39:3		
6 REP10199-7208	ELA.8.2:1	ELA.8.11:4			
7 REP10199-7209	ELA.8.10:1	ELA.8.12:4			
8 REP10199-7210	ELA.8.12:5				
9 REP10199-7211	ELA.8.15:5				
10 10199-7213	ELA.8.16:5				
11 REP10199-8834	LR.8.CS:2	ELA.8.9:1	IR.8.CS:2		
12 REP10199-8848	ELA.8.1:5				
13 REP10199-8711	ELA.8.2:4	ELA.8.14:1			
14 REP10199-8712	ELA.8.3:5				
15 REP10199-8715	ELA.8.3:1	ELA.8.7:1	ELA.8.9:3		
16 REP10199-8714	ELA.8.7:4	ELA.8.39:1			
17 10199-8851	ELA.8.7:1	ELA.8.14:2	ELA.8.39:2		
18 REP10199-8849	ELA.8.3:2	ELA.8.14:3			
19 REP10199-8847	ELA.8.3:4	ELA.8.4:1			
20 REP10199-8850	ELA.8.14:5				
21 REP10199-1262	ELA.8.37:5				
22 REP10199-1263	ELA.8.37:5				
23 REP10199-1264	ELA.8.38:8	ELA.8.39:2			
24 REP10199-1265	ELA.8.37:5				
25 REP10199-8235	ELA.8.4:5				
26 10199-8236	ELA.8.5:2	ELA.8.11:3			
27 REP10199-8291	ELA.8.11:10				
28 REP10199-8294	ELA.8.5:5				
29 REP10199-8295	ELA.8.4:1	ELA.8.10:3	ELA.8.11:1		
30 REP10199-13994	ELA.8.4:2	ELA.8.6:3			

31 10199-13997	ELA.8.4:1	ELA.8.5:1	ELA.8.11:3	
32 10199-7503	ELA.8.1:4	ELA.8.3:1		
33 10199-7521	ELA.8.1:1	ELA.8.7:3	ELA.8.39:1	
34 10199-7529	ELA.8.1:1	ELA.8.7:4		
35 10199-7530	ELA.8.2:5			
36 10199-7535	ELA.8.9:4	ELA.8.38:1		
37 10199-1282	ELA.8.38:5			
38 10199-1285	ELA.8.38:10			
39 10199-1291	ELA.8.37:5			
40 10199-1294	ELA.8.37:5			
41 CONVENTIONS10199-14170	ELA.8.36:10	ELA.8.37:10		
42 ELABORATION10199-14170	ELA.8.20:20	ELA.8.23:4		
43 ORGANIZATION10199-14170	ELA.8.20:20	ELA.8.23:20	ELA.8.36:4	ELA.8.38:12

Table 8.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV ELA 2019 Grade 8 B3

Low DOK		Matched DOK		High DOK

LR.8								
LR.8.KID								
ELA.8.1: [2]	1:(1)[2]	12:(5)[2]	32:(4)[2]	33:(1)[2]	34:(1)[2]			
ELA.8.2: [3]	4:(1)[4]	5:(1)[4]	6:(1)[4]	13:(4)[2]	35:(5)[2]			
ELA.8.3: [3]	1:(1)[2]	14:(5)[2]	15:(1)[2]	18:(2)[2]	19:(4)[2]	32:(1)[2]		
LR.8.CS: [3]	11:(2)[2]							
ELA.8.7: [3]	4:(1)[2]	5:(1)[2]	15:(1)[2]	16:(4)[2]	17:(1)[2]	33:(3)[2]	34:(4)[2]	
ELA.8.8								
ELA.8.9: [3]	11:(1)[3]	15:(3)[2]	36:(4)[2]					
LR.8.IKI								
ELA.8.13								
ELA.8.14: [3]	13:(1)[2]	17:(2)[2]	18:(3)[2]	20:(5)[3]				
IR.8								
IR.8.KID								
ELA.8.4: [2]	1:(2)[2]	2:(1)[2]	3:(1)[2]	19:(1)[2]	25:(5)[2]	29:(1)[2]	30:(2)[2]	31:(1)[2]
ELA.8.5: [3]	2:(3)[2]	26:(2)[2]	28:(5)[2]	31:(1)[2]				
ELA.8.6: [3]	3:(2)[2]	30:(3)[2]						
IR.8.CS: [3]	11:(2)[2]							
ELA.8.10: [3]	4:(3)[2]	7:(1)[4]	29:(3)[2]					
ELA.8.11: [3]	6:(4)[2]	26:(3)[2]	27:(10)[2]	29:(1)[2]	31:(3)[2]			
ELA.8.12: [3]	3:(1)[2]	7:(4)[2]	8:(5)[2]					
IR.8.IKI								
ELA.8.15: [3]	9:(5)[2]							
ELA.8.16: [3]	10:(5)[2]							
ELA.8.17								
WL.8								
WL.8.TTP								
ELA.8.20: [3]	42:(20)[3]	43:(20)[3]						

ELA.8.21								
ELA.8.22								
WL.8.PDW								
ELA.8.23: [3]	42:(4)[3]	43:(20)[3]						
ELA.8.24								
ELA.8.25								
WL.8.RBPK								
ELA.8.26								
ELA.8.27								
ELA.8.28								
WL.8.RW								
ELA.8.29								
WL.8.CSE								
ELA.8.36: [2]	1:(1)[1]	2:(1)[1]	41:(10)[1]	43:(4)[3]				
ELA.8.37: [1]	3:(1)[1]	21:(5)[1]	22:(5)[1]	24:(5)[1]	39:(5)[1]	40:(5)[1]	41:(10)[1]	
WL.8.KL								
ELA.8.38: [2]	23:(8)[2]	36:(1)[2]	37:(5)[2]	38:(10)[2]	43:(12)[3]			
WL.8.VAU								
ELA.8.39: [2]	5:(3)[2]	16:(1)[2]	17:(2)[2]	23:(2)[1]	33:(1)[2]			
ELA.8.40								
ELA.8.41								

Table 8.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 CONF WV Grade 8B3 ELA (Three Reviewers)

Low		Medium		High	
2.4		7.2		12	
1	REP10199-7201	Exact:3			
2	REP10199-7203	Exact:2	Partial:1		
3	REP10199-7205	Exact:2	Partial:1		
4	10199-7206	Exact:2	Partial:1		
5	10199-7207	Exact:3			
6	REP10199-7208	Exact:2	Partial:1		
7	REP10199-7209	Exact:2	Partial:1		
8	REP10199-7210	Exact:3			
9	REP10199-7211	Exact:2	Partial:1		
10	10199-7213	Exact:3			
11	REP10199-8834	Exact:1	Minimal:2		
12	REP10199-8848	Exact:1	Partial:2		
13	REP10199-8711	Exact:2	Partial:1		
14	REP10199-8712	Exact:2	Partial:1		
15	REP10199-8715	Partial:2	Minimal:1		
16	REP10199-8714	Exact:2	Partial:1		
17	10199-8851	Exact:2	Partial:1		
18	REP10199-8849	Exact:2	Partial:1		
19	REP10199-8847	Exact:3			
20	REP10199-8850	Partial:2	Negligible:1		
21	REP10199-1262	Exact:2	Partial:1		
22	REP10199-1263	Exact:2	Partial:1		
23	REP10199-1264	Exact:4	Partial:2		
24	REP10199-1265	Exact:2	Partial:1		
25	REP10199-8235	Exact:2	Partial:1		
26	10199-8236	Exact:1	Partial:1	Minimal:1	
27	REP10199-8291	Exact:2	Partial:4		
28	REP10199-8294	Exact:3			
29	REP10199-8295	Exact:2	Partial:1		
30	REP10199-13994	Partial:3			

31 10199-13997	Exact:1	Partial:2
32 10199-7503	Exact:3	
33 10199-7521	Exact:1	Partial:2
34 10199-7529	Exact:2	Minimal:1
35 10199-7530	Exact:3	
36 10199-7535	Exact:3	
37 10199-1282	Exact:3	
38 10199-1285	Exact:6	
39 10199-1291	Exact:3	
40 10199-1294	Exact:3	
41 CONVENTIONS10199-14170	Minimal:6	
42 ELABORATION10199-14170	Exact:12	
43 ORGANIZATION10199-14170	Partial:12	

**Information subject to nondisclosure agreements has
been redacted for public release.**

Appendix C

Reviewers' Notes

West Virginia College- and Career- Readiness Standards for ELA Grades 3-8 and Sample Test Events

2019

**Information subject to nondisclosure agreements has
been redacted for public release.**

Appendix D

Debriefing Summary Notes

West Virginia College- and Career- Readiness Standards for ELA Grades 3-8 and Sample Test Events

2019

Appendix E

DOK Definitions for Reading

2019

DOK Primer

Dr. Norman Webb originally developed the Depth of Knowledge (DOK) language system for the purpose of evaluating the relationship between the content complexity (also often referred to as “cognitive complexity”) of curriculum standards and of corresponding assessment items. Over the years, the use of DOK has extended far beyond the original context of evaluating the alignment of statewide summative assessments. Now used extensively throughout U.S. school districts, by state departments of education, assessment developers, educational publishers, and others, DOK is applied toward informing alignment between and among all areas of the education system, not just the relationship between standards and assessments. DOK is a tool that allows educators to communicate effectively, consistently, and efficiently about the content complexity of standards, learning objectives, tasks, prompts, questions, etc.

What DOK is	What DOK is NOT
- DOK is an evaluative tool used for content analysis. Specifically, DOK is a language system that can be used to differentiate between and among different levels of complexity of student engagement expressed within components of educational materials/systems. - DOK can be used to interpret standards, learning objectives, tasks, prompts, questions, etc. - DOK is a reflective lens used to foster intentionality in teachers’ practices, to help ensure that the complexity of expected learning outcomes are clearly understood, that (formative/summative/etc.) assessments provide opportunities to make reasonable inferences about attainment of learning outcomes, and that appropriate educational opportunities are provided to allow students to engage at the level(s) of complexity intended. - DOK is a conceptualization of complexity, as differentiated from difficulty.	- DOK is not used to evaluate text complexity, topic complexity (e.g. how complex is photosynthesis), phenomenon complexity (NGSS), or overall complexity of, for example, an entire lesson or unit. - DOK is not a rubric. - DOK is not a protocol. - DOK is not a type of question (The idea of “DOK questions” is not consistent with the intent of DOK.) - DOK is not a measurement of “how deeply” an individual is engaging with a topic. - DOK is not hierarchical or progressive (i.e. it does not reflect any sort of learning progression from low → high complexity). - DOK is not a value judgment and does not reflect importance. (In other words, there is no idea inherent to DOK that any level of DOK is “better” than any other. The standards, by definition, specify what is important.)

Selected Examples of how DOK Is Commonly Used

- By individual K-12 teachers and within school Professional Learning Communities (PLCs) to understand the meaning of academic standards – e.g. the complexity at which students are expected to engage with a concept or idea.
- By K-12 teachers and associated C&I staff to inform development of lesson plans, unit designs, formative/summative classroom and district assessments, and other materials.
- By K-12 teachers and associated C&I staff to inform selection of assessment items from item bank products for which a district has purchased access.
- To communicate expectations to item writers / content developers – e.g. expected distribution of DOK levels of items on an assessment.
- In specifications for assessments via state RFPs.
- By large-scale state department of education efforts to provide educators with tools for goals of school improvement, assessment development, and other endeavors.
- As one component of alignment analyses of statewide summative assessments with standards (as specified by “depth and breadth” requirement of ESSA).

Point of clarification: We use DOK to do *content analyses* of the “content” of standards, tasks, instructional prompts, questions, etc. The “content” of a standard can be thought of as “the entirety of the standard.” Hence, when we use the term “content complexity” we are referring to the “content” within a learning expectation (standard) or a task, prompt, question, etc. We can analyze the complexity of the “content” of these expectations/tasks/etc. (everything that is within) that we read on paper (or on screen). The term “content complexity” does NOT refer to the complexity of a topic (e.g. photosynthesis), text, or context, etc.



Reading DOK Definitions

DOK 1

DOK 1 involves reading text orally and with basic comprehension, decoding words, blending phonemes, receiving and reciting facts, demonstrating letter and word knowledge, and recognizing text features and common spelling patterns. DOK 1 also includes receiving or reciting facts acquired by processing text as well as reading orally without the analysis of text. Very basic comprehension of a text gained from knowledge of vocabulary and explicit structure of the text is at this category. Tasks require only a shallow understanding of the text presented and often consist of verbatim recall from text, slight paraphrasing of specific details from the text, or simple understanding of a single word or phrase. Younger students who answer direct questions about features stated explicitly in the text are performing at this category. Applying phonics and word analysis skills in decoding words are also DOK 1 tasks. Some examples that represent, but do not constitute all of, DOK 1 performance are:

- Support ideas with reference to verbatim (or only slightly paraphrased) details from the text.
- Use a dictionary to find the meanings of words.
- Recognize figurative language in a reading passage

Webb, N. L. *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states*. A study of the State Collaborative on Assessment & Student Standards (SCASS) Technical Issues in Large-Scale Assessment (TILSA). Washington, D. C.: Council of Chief State School Officers, December 2002. Revised in 2014 by Norman Webb and Sara Christopherson.

WebbAlign offers alignment studies and DOK professional development training. Please contact WCEPS at webbalign@wceps.org or 1-877-249-4211 for more information.

DOK 2

DOK 2 involves drawing meaning from text by using organizational structure, evidence, and context; summarizing main ideas, character traits, plots, themes, and figurative use of words; following cause-effect sequences and multiple ideas through a text; distinguishing among hypotheses and givens as well as fact from opinion; and explaining differences among genres (poetry, expository materials, fiction, etc.). DOK 2 requires the engagement of some mental processing beyond recalling or reproducing a response; it requires both comprehension and subsequent processing of text or portions of text. Inter-sentence analysis or inference is required. DOK 2 tasks may require use of specific information from the text to explain given events and ideas. At this level, reading concepts (e.g. making inferences or predictions) are generally applied for purposeful reading. Multiple features of the text are processed to gain a deeper understanding of the text such as organizing in a time sequence, outlining, comparing fact from opinion, and using graphic aides. Deciphering main ideas supported by key details or drawing on details to describe a feature in a story are stressed. Younger students conveying important points from a story fit under this category. DOK 2 ideas, in general, apply the skills and concepts that constitute DOK 1. However, DOK 2 activities involve closer understanding of text, possibly through paraphrasing, such as putting in one's own words both the question and response to an assessment item. Some examples that represent, but do not constitute all of, DOK 2 performance include:

- Use context cues to identify the meaning of unfamiliar words, phrases, and expressions that could otherwise have multiple meanings.
- Predict a logical outcome based on information in a reading selection.
- Identify and summarize the major events in a narrative.

DOK 3

DOK 3 involves conducting analyses of the text to make inferences on author's purpose and use of textual features (e.g. literary devices to support and convey the main message); engaging in critical reading to attest to the credibility of the message, the internal logic, and implied values, attitudes, and biases; and going beyond the text by comparing features and meaning with other texts, considering the impact of the time period and other conditions when the text was written, and raising valid alternative hypotheses and conclusions to those presented in the text. At DOK 3 deep knowledge becomes a greater focus. Students are encouraged to go beyond the text; however, they are still required to show understanding of the ideas in the text. Students may be encouraged to explain, generalize, or connect ideas while applying reasoning and planning. Students must be able to support their thinking. Younger students who provide some valid evidence for their breakdown of a story into meaningful parts are performing at this category. Tasks at a Category 3 may involve abstract theme identification, inference across an entire passage with multiple paragraphs, or students' application of prior knowledge. Activities may also involve identifying more abstract connections between texts. Some examples that represent, but do not constitute all of, DOK 3 performance include:

- Explain or recognize how the author's purpose affects the interpretation of a reading selection.
- Summarize information from multiple sources to address a specific topic.
- Analyze and describe the characteristics of various types of literature.

DOK 4

DOK 4 involves at least as complex content as in the previous category, but also requires working on a task over an extended period of time such as when conducting a research project over weeks. The extended time that accompanies this type of activity allows for creation of original work and requires metacognitive awareness that typically increases the complexity of a DOK 4 task overall, in comparison with DOK 3 activities. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require the application of significant conceptual understanding and higher-order thinking.

DOK 4 activities may have students take information from multiple passages and texts to find supporting evidence and counter points for developing an argument or reaching conclusions or could involve creating an original thesis on a topic based on information drawn from relevant references. For younger students, an extended period of time could be multiple days for reaching conclusions from reading a number of texts. Students take information from a multiple of passages and are asked to apply this information to a new task. They may also be asked to develop hypotheses and perform complex analyses of the connections among texts requiring work over an extended period of time. Some examples that represent, but do not constitute all of, DOK 4 performance are:

- Analyze and synthesize information from multiple sources.
- Examine and explain alternative perspectives across a variety of sources.
- Describe and illustrate how common themes are found across texts from different cultures.

General Guidelines for Assigning DOK:

- The DOK definitions can be applied to reading standards, tasks, or activities.
- Consider the complexity of the reading demands, not the difficulty for students.
- Consider the experience (prior knowledge) and grade-level expectations of a typical student.
- Do not rely on verbs (describe, explain, evaluate, etc.). Instead, consider the content complexity required for an adequate response.
- For multiple-choice assessment items, consider the item as a whole—including distractors—to judge complexity.
- An expectation or item that is confusing due to error or wording does not reflect increased content complexity—it simply means the statement needs revisions.
- The reading DOK levels were originally based on Valencia and Wixson (2000, pp. 909-935).

Appendix F

Coding Instructions Provided to Panelists

2019

Coding How-To/Reminders for Reviewers

WV Alignment Study, June, 2019

STEP ONE: INDEPENDENT ANALYSIS

- Identify what knowledge is **necessary and sufficient** to answer the item correctly. Code item to corresponding standard, choosing the most specific standard possible.
- ONLY if the **necessary and sufficient** knowledge is expressed in more than one standard, code a primary standard and one or two secondary standard(s).
- The content of an item may not correspond to a particular standard but still fit in that “generic” level. Code to the closest generic standard and add a reason why the item was coded to that generic.
- The standards apply to all students; think about what the “typical” grade-level student will need to do to respond to a prompt.
- Consider the full scope of standards.
- Always refer to the DOK definitions to ensure consistent coding.
- Write a reason for any item that contains a Source-of-Challenge issue. Use only in fairly extreme situations when items are technically flawed – i.e. could be answered correctly for the wrong reason or incorrectly for the wrong reason.
- Other comments about strong items, weak items, or items that are perceived as “good” or “bad” should go into the Notes box.
- Write brief comments on any item, but don’t slow down the coding process.
- Record your standard and DOK assignments on both laptops...just in case you get off-number.
- SAVE, SAVE, SAVE!

Debriefing Questions:

- Responses to the debriefing questions allow study directors to obtain a greater sense of the qualitative topics and issues associated with that particular grade-level analysis.
- Do not repeat any information already entered in DOK/standard coding.
- Add feedback or other information not captured in the DOK/standard data.

STEP TWO: CONFIRMATORY ANALYSIS

For each item, consider the internally coded standard(s).

To what extent does the item assess the content (expectations) within the internally coded standard(s)?

- **EXACTLY** (Note that the item does NOT need to assess every aspect of a standard, but it needs to be a direct (“exact”) measurement of a central aspect of the standard. A correct response to the item allows for a direct inference about student knowledge/skills/abilities as expressed in the standard.)
- **PARTIALLY** (The item somewhat targets the expectations within the standard and it can be considered a majority match. A correct response to the item allows for some inference about student knowledge/skills/abilities as expressed in the standard.)
- **MINIMALLY** (The item only very minimally targets the expectations within the standard – and it can be considered only a minority match. A correct response to the item allows for very little or very indirect inference about student knowledge/skills/abilities as expressed in the standard.)
- **NOT AT ALL** (The item does not assess the expectations within the standard. No inference can be made about student knowledge/skills/abilities as expressed in the standard based on a correct response.)

For each item, consider the internally assigned DOK.

- **If you agree with the internally assigned DOK**, enter that value (1, 2, or 3) into the drop-down menu on the WATv2 for that item.
- **If you do not agree with the internally assigned DOK**, enter the DOK that you assigned to the item.



Appendix A

Group Consensus DOK Values for West Virginia College- and Career- Readiness Standards for Mathematics Grades 3-8

2019

	West Virginia College and Career Readiness Standards for Grade 3 Mathematics	
Standard	Description	DOK
OA.M.3	Operations and Algebraic Thinking (OA)	
OA.M.3A	Represent and solve problems involving multiplication and division.	
M.3.1	Interpret products of whole numbers, e.g., interpret 5×7 as the total number of objects in 5 groups of 7 objects each. <i>For example, describe a context in which a total number of objects can be expressed as 5×7.</i>	2
M.3.2	Interpret whole-number quotients of whole numbers, e.g., interpret $56 \div 8$ as the number of objects in each share when 56 objects are partitioned equally into 8 shares, or as a number of shares when 56 objects are partitioned into equal shares of 8 objects each. <i>For example, describe a context in which a number of shares or a number of groups can be expressed as $56 \div 8$.</i>	2
M.3.3	Use multiplication and division within 100 to solve word problems in situations involving equal groups, arrays, and measurement quantities, e.g., by using drawings and equations with a symbol for the unknown number to represent the problem.	2
M.3.4	Determine the unknown whole number in a multiplication or division equation relating three whole numbers. <i>For example, determine the unknown number that makes the equation true in each of the equations $8 \times ? = 48$, $5 = ? \div 3$, $6 \times 6 = ?$.</i>	1
OA.M.3B	Understand properties of multiplication and the relationship between multiplication and division.	
M.3.5	Apply properties of operations as strategies to multiply and divide. (e.g., <i>If $6 \times 4 = 24$ is known, then $4 \times 6 = 24$ is also known: Commutative property of multiplication. $3 \times 5 \times 2$ can be found by $3 \times 5 = 15$, then $15 \times 2 = 30$, or by $5 \times 2 = 10$, then $3 \times 10 = 30$: Associative property of multiplication. Knowing that $8 \times 5 = 40$ and $8 \times 2 = 16$, one can find 8×7 as $8 \times (5 + 2) = (8 \times 5) + (8 \times 2) = 40 + 16 = 56$: Distributive property.</i>	2
M.3.6	Understand division as an unknown-factor problem. (e.g., <i>find $32 \div 8$ by finding the number that makes 32 when multiplied by 8</i>).	2

OA.M.3C	Multiply and divide within 100.	
M.3.7	Learn multiplication tables (facts) with speed and memory in order to fluently multiply and divide within 100, using strategies such as the relationship between multiplication and division (e.g., knowing that $8 \times 5 = 40$, one knows that $40 \div 5 = 8$) or properties of operations by the end of Grade 3.	1
OA.M.3D	Solve problems involving the four operations, and identify and explain patterns in arithmetic.	
M.3.8	Solve two-step word problems using the four operations. Represent these problems using equations with a letter standing for the unknown quantity. Assess the reasonableness of answers using mental computation and estimation strategies including rounding.	2
M.3.9	Identify arithmetic patterns (including patterns in the addition table or multiplication table), and explain those using properties of operations (e.g., observe that 4 times a number is always even, and explain why 4 times a number can be decomposed into two equal addends).	2
NBT.NF.M.3	Number and Operations in Base Ten (NBT) and Fractions (NF)	
NBT.NF.M.3A	Use place value understanding and properties of operations to perform multi-digit arithmetic.	
M.3.10	Use place value understanding to round whole numbers to the nearest 10 or 100.	1
M.3.11	Fluently add and subtract within 1000 using strategies and algorithms based on place value, properties of operations, and/or the relationship between addition and subtraction.	1
M.3.12	Multiply one-digit whole numbers by multiples of 10 in the range 10–90 (e.g., 9×80 , 5×60) using strategies based on place value and properties of operations.	1

NBT.NF.M.3B	Develop understanding of fractions as numbers.	
M.3.13	Understand a fraction $1/b$ as the quantity formed by 1 part when a whole is partitioned into b equal parts; understand a fraction a/b as the quantity formed by a parts of size $1/b$.	1
M.3.14	Understand a fraction as a number on the number line and represent fractions on a number line diagram. a. Represent a fraction $1/b$ on a number line diagram by defining the interval from 0 to 1 as the whole and partitioning it into b equal parts. Recognize that each part has size $1/b$ and that the endpoint of the part based at 0 locates the number $1/b$ on the number line. (e.g., Given that b parts is 4 parts, then $1/b$ represents $1/4$. Students partition the number line into fourths and locate $1/4$ on the number line.) b. Represent a fraction a/b on a number line diagram by marking off a lengths $1/b$ from 0. Recognize that the resulting interval has size a/b and that its endpoint locates the number a/b on the number line. (e.g., Given that a/b represents $3/4$ or $6/4$, students partition the number line into fourths and represent these fractions accurately on the same number line; students extend the number line to include the number of wholes required for the given fractions.)	2
M.3.15	Explain equivalence of fractions in special cases, and compare fractions by reasoning about their size. a. Understand two fractions as equivalent (equal) if they are the same size, or the same point on a number line. b. Recognize and generate simple equivalent fractions, e.g., $1/2 = 2/4$, $4/6 = 2/3$. Explain why the fractions are equivalent, e.g., by using a visual fraction model. c. Express whole numbers as fractions, and recognize fractions that are equivalent to whole numbers. <i>Examples: Express 3 in the form $3 = 3/1$; recognize that $6/1 = 6$; locate $4/4$ and 1 at the same point of a number line diagram.</i> d. Compare two fractions with the same numerator or the same denominator by reasoning about their size. Recognize that comparisons are valid only when the two fractions refer to the same whole. Record the results of comparisons with the symbols $>$, $=$, or $<$, and justify the conclusions, (e.g., by using a visual fraction model.)	2

MD.G.M.3	Measurement and Data (MD) and Geometry (G)	
MD.G.M.3A	Solve problems involving measurement and estimation of intervals of time, liquid volumes, and masses of objects.	
M.3.16	Tell and write time to the nearest minute and measure time intervals in minutes. Solve word problems involving addition and subtraction of time intervals in minutes (e.g., by representing the problem on a number line diagram.)	2
M.3.17	Measure and estimate liquid volumes and masses of objects using standard units of grams (g), kilograms (kg), and liters (l). Add, subtract, multiply, or divide to solve one-step word problems involving masses or volumes that are given in the same units, (e.g., by using drawings (such as a beaker with a measurement scale) to represent the problem.	2
MD.G.M.3B	Represent and interpret data.	
M.3.18	Draw a scaled picture graph and a scaled bar graph to represent a data set with several categories. Solve one- and two-step “how many more” and “how many less” problems using information presented in scaled bar graphs. (e.g., <i>draw a bar graph in which each square in the bar graph might represent 5 pets.</i>)	2
M.3.19	Generate measurement data by measuring lengths using rulers marked with halves and fourths of an inch. Show the data by making a line plot, where the horizontal scale is marked off in appropriate units— whole numbers, halves, or quarters.	2
MD.G.M.3C	Geometric measurement: understand concepts of area and relate area to multiplication and to addition.	
M.3.20	Recognize area as an attribute of plane figures and understand concepts of area measurement. a. A square with side length 1 unit, called “a unit square,” is said to have “one square unit” of area, and can be used to measure area. b. A plane figure which can be covered without gaps or overlaps by b unit squares is said to have an area of b square units.	1
M.3.21	Measure areas by counting unit squares (square cm, square m, square in, square ft, and improvised units).	1

M.3.22	Relate area to the operations of multiplication and addition. a. Find the area of a rectangle with whole-number side lengths by tiling it, and show that the area is the same as would be found by multiplying the side lengths. b. Multiply side lengths to find areas of rectangles with whole-number side lengths in the context of solving real world and mathematical problems, and represent whole-number products as rectangular areas in mathematical reasoning. c. Use tiling to show in a concrete case that the area of a rectangle with whole-number side lengths a and $b + c$ is the sum of $a \times b$ and $a \times c$. Use area models to represent the distributive property in mathematical reasoning. d. Recognize area as additive. Find areas of rectilinear figures by decomposing them into non-overlapping rectangles and adding the areas of the non-overlapping parts, applying this technique to solve real world problems.	2
MD.G.M.3D	Geometric measurement: Recognize perimeter as an attribute of plane figures and distinguish between linear and area measures.	
M.3.23	Solve real world and mathematical problems involving perimeters of polygons, including finding the perimeter given the side lengths, finding an unknown side length, and exhibiting rectangles with the same perimeter and different areas or with the same area and different perimeters.	2
MD.G.M.3E	Reason with shapes and their attributes.	
M.3.24	Understand that shapes in different categories (e.g., rhombuses, rectangles, circles, and others) may share attributes (e.g., having four sides), and that the shared attributes can define a larger category (e.g., quadrilaterals). Recognize rhombuses, rectangles, and squares as examples of quadrilaterals, and draw examples of quadrilaterals that do not belong to any of these subcategories.	2
M.3.25	Partition shapes into parts with equal areas. Express the area of each part as a unit fraction of the whole. For example, partition a shape into 4 parts with equal area, and describe the area of each part as $\frac{1}{4}$ or the area of the shape.	2

	West Virginia College and Career Ready Standards for Grade 4 Mathematics	
Standard	Description	DOK
OA.M.4	Operations and Algebraic Thinking (OA)	
OA.M.4A	Use the four operations with whole numbers to solve problems.	
M.4.1	Interpret a multiplication equation as a comparison, e.g., interpret $35 = 5 \times 7$ as a statement that 35 is 5 times as many as 7 and 7 times as many as 5. Represent verbal statements of multiplicative comparisons as multiplication equations.	2
M.4.2	Multiply or divide to solve word problems involving multiplicative comparison, e.g., by using drawings and equations with a symbol for the unknown number to represent the problem, and distinguish multiplicative comparison from additive comparison.	2
M.4.3	Solve multi-step word problems posed with whole numbers and having whole-number answers using the four operations, including problems in which remainders must be interpreted. Represent these problems using equations with a letter standing for the unknown quantity. Assess the reasonableness of answers using mental computation and estimation strategies including rounding.	2
OA.M.4B	Gain familiarity with factors and multiples.	
M.4.4	Find all factor pairs for a whole number in the range 1–100. Recognize that a whole number is a multiple of each of its factors. Determine whether a given whole number in the range 1–100 is a multiple of a given one-digit number. Determine whether a given whole number in the range 1–100 is prime or composite.	1
OA.M.4C	Generate and analyze patterns.	
M.4.5	Generate a number or shape pattern that follows a given rule. Identify apparent features of the pattern that were not explicit in the rule itself. (e.g., given the rule “Add 3” and the starting number 1, generate terms in the resulting sequence and observe that the terms appear to alternate between odd and even numbers. Explain informally why the numbers will continue to alternate in this way.)	2

NBT.NF.M.4	Number and Operations in Base Ten (NBT) and Fractions (NF)	
NBT.NF.M.4A	Generalize place value understanding for multi-digit whole numbers.	
M.4.6	Recognize that in a multi-digit whole number, a digit in one place represents ten times what it represents in the place to its right. <i>(For example, recognize that $700 \div 70 = 10$ by applying concepts of place value and division.)</i>	1
M.4.7	Read and write multi-digit whole numbers using base-ten numerals, number names, and expanded form. Compare two multi-digit numbers based on meanings of the digits in each place, using $>$, $=$, and $<$ symbols to record the results of comparisons.	1
M.4.8	Use place value understanding to round multi-digit whole numbers to any place.	1
NBT.NF.M.4B	Use place value understanding and properties of operations to perform multi-digit arithmetic.	
M.4.9	Fluently add and subtract multi-digit whole numbers using the standard algorithm.	1
M.4.10	Multiply a whole number of up to four digits by a one-digit whole number, and multiply two two-digit numbers, using strategies based on place value and the properties of operations. Illustrate and explain the calculation by using equations, rectangular arrays, and/or area models.	2
M.4.11	Find whole-number quotients and remainders with up to four-digit dividends and one-digit divisors, using strategies based on place value, the properties of operations, and/or the relationship between multiplication and division. Illustrate and explain the calculation by using equations, rectangular arrays, and/or area models.	2
NBT.NF.M.4C	Extend understanding of fraction equivalence and ordering.	
M.4.12	Explain why a fraction a/b is equivalent to a fraction $(n \times a)/(n \times b)$ by using visual fraction models, with attention to how the number and size of the parts differ even though the two fractions themselves are the same size. Use this principle to recognize and generate equivalent fractions.	2
M.4.13	Compare two fractions with different numerators and different denominators, e.g., by creating common denominators or numerators, or by comparing to a benchmark fraction such as $1/2$. Recognize that comparisons are valid only when the two fractions refer to the same whole. Record the results of comparisons with symbols $>$, $=$, or $<$, and justify the conclusions, e.g., by using a visual fraction model.	2

NBT.NF.M.4D	Build fractions from unit fractions from by applying and extending previous understandings of operations on whole numbers.	
M.4.14	Understand a fraction a/b with $a > 1$ as a sum of fractions $1/b$. a. Understand addition and subtraction of fractions as joining and separating parts referring to the same whole. b. Decompose a fraction into a sum of fractions with the same denominator in more than one way, recording each decomposition by an equation. Justify decompositions, e.g., by using a visual fraction model. Examples: $3/8 = 1/8 + 1/8 + 1/8$; $3/8 = 1/8 + 2/8$; $2\ 1/8 = 1 + 1 + 1/8 = 8/8 + 8/8 + 1/8$. c. Add and subtract mixed numbers with like denominators by replacing each mixed number with an equivalent fraction, and/or by using properties of operations and the relationship between addition and subtraction. d. Solve word problems involving addition and subtraction of fractions referring to the same whole and having like denominators by using visual fraction models and equations to represent the problem.	2
M.4.15	Apply and extend previous understandings of multiplication to multiply a fraction by a whole number. a. Understand a fraction a/b as a multiple of $1/b$. For example, use a visual fraction model to represent $5/4$ as the product $5 \times (1/4)$, recording the conclusion by the equation $5/4 = 5 \times (1/4)$. b. Understand a multiple of a/b as a multiple of $1/b$, and use this understanding to multiply a fraction by a whole number. For example, use a visual fraction model to express $3 \times (2/5)$ as $6 \times (1/5)$, recognizing this product as $6/5$. (In general, $n \times (a/b) = (n \times a)/b$.) c. Solve word problems involving multiplication of a fraction by a whole number by using visual fraction models and equations to represent the problem. For example, if each person at a party will eat $3/8$ of a pound of roast beef, and there will be 5 people at the party, how many pounds of roast beef will be needed? Between what two whole numbers does your answer to lie?	2
NBT.NF.M.4E	Understand decimal notation for fractions, and compare decimal fractions.	
M.4.16	Express a fraction with denominator 10 as an equivalent fraction with denominator 100, and use this technique to add two fractions with respective denominators 10 and 100. For example, express $3/10$ as $30/100$, and add $3/10 + 4/100 = 34/100$.	1
M.4.17	Use decimal notation for fractions with denominators 10 or 100. For example, rewrite 0.62 as $62/100$; describe a length as 0.62 meters; locate 0.62 on a number line diagram.	1
M.4.18	Compare two decimals to hundredths by reasoning about their size. Recognize that comparisons are valid only when the two decimals refer to the same whole. Record the results of comparisons with the symbols $>$, $=$, or $<$, and justify the conclusions by using a visual model.	2

MD.G.M.4	Measurement and Data (MD) and Geometry (G)	
MD.G.M.4A	Solve problems involving measurement and conversion of measurements from a larger unit to a smaller unit.	
M.4.19	Know relative sizes of measurement units within a system of units, including the metric system (km, m, cm; kg, g; l, ml), the standard system (lb, oz), and time (hr, min, sec.). Within a single system of measurement, express measurements in a larger unit in terms of a smaller unit. Record measurement equivalents in a two-column table. For example, know that 1 ft is 12 times as long as 1 in. Express the length of a 4 ft snake as 48 in. Generate a conversion table for feet and inches listing the number pairs (1, 12), (2, 24), (3, 36),...	1
M. 4.20	Use the four operations to solve word problems involving distances, intervals of time, liquid volumes, masses of objects, and money, including problems involving simple fractions or decimals and problems that require expressing measurements given in a larger unit in terms of a smaller unit. Represent measurement quantities using diagrams such as number line diagrams that feature a measurement scale.	2
M.4.21	Apply the area and perimeter formulas for rectangles in real world and mathematical problems by viewing the area formula as a multiplication equation with an unknown factor. For example, find the width of a rectangular room given the area of the flooring and the length.	2
MD.G.M.4B	Represent and interpret data.	
M.4.22	Make a line plot to display a data set of measurements in fractions of a unit ($\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$). Solve problems involving addition and subtraction of fractions by using information presented in line plots. For example, from a line plot find and interpret the difference in length between the longest and shortest specimens in an insect collection.	2
MD.G.M.4C	Geometric measurement: understand concepts of angle and measure angles.	
M.4.23	Recognize angles as geometric shapes that are formed wherever two rays share a common endpoint, and understand concepts of angle measurement: - An angle is measured with reference to a circle with its center at the common endpoint of the rays, by considering the fraction of the circular arc between the points where the two rays intersect the circle. An angle that turns through $\frac{1}{360}$ of a circle is called a “one-degree angle,” and can be used to measure angles. - An angle that turns through b one-degree angles is said to have an angle measure of b degrees.	1
M.4.24	Measure angles in whole-number degrees using a protractor. Sketch angles of specified measure.	1

M.4.25	Recognize angle measure as additive. When an angle is decomposed into nonoverlapping parts, the angle measure of the whole is the sum of the angle measures of the parts. Solve addition and subtraction problems to find unknown angles on a diagram in real world and mathematical problems, (e.g., by using an equation with a symbol for the unknown angle measure).	2
MD.G.M.4D	Draw and identify lines and angles, and classify shapes by properties of their lines and angles.	
M.4.26	Draw points, lines, line segments, rays, angles (right, acute, obtuse), and perpendicular and parallel lines. Identify these in two-dimensional figures.	1
M.4.27	Classify two-dimensional figures based on the presence or absence of parallel or perpendicular lines, or the presence or absence of angles of a specified size. Recognize right triangles as a category, and identify right triangles.	1
M.4.28	Recognize a line of symmetry for a two-dimensional figure as a line across the figure such that the figure can be folded along the line into matching parts. Identify line-symmetric figures and draw lines of symmetry.	1

	West Virginia College and Career Ready Standards for Grade 5 Mathematics	
Standard	Description	DOK
OA.M.5	Operations and Algebraic Thinking (OA)	
OA.M.5A	Write and interpret numerical expressions.	
M.5.1	Use parentheses, brackets or braces in numerical expressions and evaluate expressions with these symbols.	1
M.5.2	Write simple expressions that record calculations with numbers, and interpret numerical expressions without evaluating them. For example, express the calculation “add 8 and 7, then multiply by 2” as $2 \times (8 + 7)$. Recognize that $3 \times (18932 + 921)$ is three times as large as $18932 + 921$, without having to calculate the indicated sum or product.	2
OA.M.5B	Analyze patterns and relationships.	
M.5.3	Generate two numerical patterns using two given rules. Identify apparent relationships between corresponding terms. Form ordered pairs consisting of corresponding terms from the two patterns, and graph the ordered pairs on a coordinate plane. For example, given the rule “Add 3” and the starting number 0, and given the rule “Add 6” and the starting number 0, generate terms in the resulting sequences, and observe that the terms in one sequence are twice the corresponding terms in the other sequence. Explain informally why this is so.	2
NBT.NF.M.5	Number and Operations in Base Ten (NBT) and Fractions (NF)	
NBT.NF.M.5A	Understand the place value system.	
M.5.4	Recognize that in a multi-digit number, a digit in one place represents 10 times as much as it represents in the place to its right and $1/10$ of what it represents in the place to its left.	2
M.5.5	Explain patterns in the number of zeros of the product when multiplying a number by powers of 10, and explain patterns in the placement of the decimal point when a decimal is multiplied or divided by a power of 10. Use whole-number exponents to denote powers of 10.	2

M.5.6	Read, write, and compare decimals to thousandths. a. Read and write decimals to thousandths using base-ten numerals, number names, and expanded form, e.g., $347.392 = 3 \times 100 + 4 \times 10 + 7 \times 1 + 3 \times (1/10) + 9 \times (1/100) + 2 \times (1/1000)$. b. Compare two decimals to thousandths based on meanings of the digits in each place, using $>$, $=$, and $<$ symbols to record the results of comparisons.	1
M.5.7	Use place value understanding to round decimals to any place.	1
NBT.NF.M.5B	Perform operations with multi-digit whole numbers and with decimals to hundredths.	
M.5.8	Fluently multiply multi-digit whole numbers using the standard algorithm.	1
M.5.9	Find whole-number quotients of whole numbers with up to four-digit dividends and two-digit divisors, using strategies based on place value, the properties of operations, and/or the relationship between multiplication and division. Illustrate and explain the calculation by using equations, rectangular arrays, and/or area models.	2
M.5.10	Add, subtract, multiply, and divide decimals to hundredths, using concrete models or drawings and strategies based on place value, properties of operations, and/or the relationship between addition and subtraction; relate the strategy to a written method and explain the reasoning used.	2
NBT.NF.M.5C	Use equivalent fractions as a strategy to add and subtract fractions.	
M.5.11	Add and subtract fractions with unlike denominators, including mixed numbers, by replacing given fractions with equivalent fractions in such a way as to produce an equivalent sum or difference of fractions with like denominators. For example, $2/3 + 5/4 = 8/12 + 15/12 = 23/12$.	1
M.5.12	Solve word problems involving addition and subtraction of fractions referring to the same whole, including cases of unlike denominators, e.g., by using visual fraction models or equations to represent the problem. Use benchmark fractions and number sense of fractions to estimate mentally and assess the reasonableness of answers. For example, recognize an incorrect result $2/5 + 1/2 = 3/7$, by observing that $3/7 < 1/2$.	2

NBT.NF.M.5D	Apply and extend previous understandings of multiplication and division to multiply and divide fractions.	
M.5.13	Interpret a fraction as division of the numerator by the denominator ($a/b = a \div b$). Solve word problems involving division of whole numbers leading to answers in the form of fractions or mixed numbers by using visual fraction models or equations to represent the problem. For example, interpret $3/4$ as the result of dividing 3 by 4, noting that $3/4$ multiplied by 4 equals 3, and that when 3 wholes are shared equally among 4 people each person has a share of size $3/4$. If 9 people want to share a 50-pound sack of rice equally by weight, how many pounds of rice should each person get? Between what two whole numbers does your answer lie?	2
M.5.14	Apply and extend previous understandings of multiplication to multiply a fraction or whole number by a fraction. a. Interpret the product $(a/b) \times q$ as a parts of a partition of q into b equal parts; equivalently, as the result of a sequence of operations $a \times q \div b$. For example, use a visual fraction model to show $(2/3) \times 4 = 8/3$, and create a story context for this equation. Do the same with $(2/3) \times (4/5) = 8/15$. b. Find the area of a rectangle with fractional side lengths by tiling it with unit squares of the appropriate unit fraction side lengths, and show that the area is the same as would be found by multiplying the side lengths. Multiply fractional side lengths to find areas of rectangles, and represent fraction products as rectangular areas.	2
M.5.15	Interpret multiplication as scaling (resizing), by: a. Comparing the size of a product to the size of one factor on the basis of the size of the other factor, without performing the indicated multiplication. b. Explaining why multiplying a given number by a fraction greater than 1 results in a product greater than the given number (recognizing multiplication by whole numbers greater than 1 as a familiar case); explaining why multiplying a given number by a fraction less than 1 results in a product smaller than the given number; and relating the principle of fraction equivalence $a/b = (n \times a)/(n \times b)$ to the effect of multiplying a/b by 1.	2
M.5.16	Solve real world problems involving multiplication of fractions and mixed numbers by using visual fraction models or equations to represent the problem.	2

M.5.17	Apply and extend previous understandings of division to divide unit fractions by whole numbers and whole numbers by unit fractions. a. Interpret division of a unit fraction by a non-zero whole number, and compute such quotients. (e.g. create a story context for $(1/3) \div 4$, and use a visual fraction model to show the quotient. Use the relationship between multiplication and division to explain that $(1/3) \div 4 = 1/12$ because $(1/12) \times 4 = 1/3$.) b. Interpret division of a whole number by a unit fraction, and compute such quotients. (e.g. create a story context for $4 \div (1/5)$, and use a visual fraction model to show the quotient. Use the relationship between multiplication and division to explain that $4 \div (1/5) = 20$ because $20 \times (1/5) = 4$.) c. Solve real world problems involving division of unit fractions by non-zero whole numbers and division of whole numbers by unit fractions, e.g., by using visual fraction models and equations to represent the problem. (e.g. how much chocolate will each person get if 3 people share $1/2$ lb of chocolate equally? How many $1/3$-cup servings are in 2 cups of raisins?)	2
MD.G.M.5	Measurement and Data (MD) and Geometry (G)	
MD.G.M.5A	Convert like measurement units within a given measurement system.	
M.5.18	Convert among different-sized standard measurement units within a given measurement system (e.g., convert 5 cm to 0.05 m), and use these conversions in solving multi-step, real world problems.	2
MD.G.M.5B	Represent and interpret data.	
M.5.19	Make a line plot to display a data set of measurements in fractions of a unit ($1/2$, $1/4$, $1/8$). Use operations on fractions for this grade to solve problems involving information presented in line plots. (e.g., given different measurements of liquid in identical beakers, find the amount of liquid each beaker would contain if the total amount in all the beakers were redistributed equally.)	2
MD.G.M.5C	Geometric measurement: understand concepts of volume and relate volume to multiplication and to addition.	
M.5.20	Recognize volume as an attribute of solid figures and understand concepts of volume measurement. a. A cube with side length 1 unit, called a “unit cube,” is said to have “one cubic unit” of volume and can be used to measure volume. b. A solid figure which can be packed without gaps or overlaps using b unit cubes is said to have a volume of b cubic units.	1
M.5.21	Measure volumes by counting unit cubes, using cubic cm, cubic in, cubic ft, and improvised units.	1

M.5.22	Relate volume to the operations of multiplication and addition and solve real world and mathematical problems involving volume. a. Find the volume of a right rectangular prism with whole-number side lengths by packing it with unit cubes, and show that the volume is the same as would be found by multiplying the edge lengths, equivalently by multiplying the height by the area of the base. Represent threefold whole-number products as volumes,(e.g., to represent the associative property of multiplication). b. Apply the formulas $V = l \times w \times h$ and $V = b \times h$ for rectangular prisms to find volumes of right rectangular prisms with whole-number edge lengths in the context of solving real world and mathematical problems. c. Recognize volume as additive. Find volumes of solid figures composed of two nonoverlapping right rectangular prisms by adding the volumes of the non-overlapping parts, applying this technique to solve real world problems.	2
MD.G.M.5D	Graph points on the coordinate plane to solve real-world and mathematical problems.	
M.5.23	Use a pair of perpendicular number lines, called axes, to define a coordinate system, with the intersection of the lines (the origin) arranged to coincide with the 0 on each line and a given point in the plane located by using an ordered pair of numbers, called its coordinates. Understand that the first number indicates how far to travel from the origin in the direction of one axis, and the second number indicates how far to travel in the direction of the second axis, with the convention that the names of the two axes and the coordinates correspond (e.g., x-axis and x-coordinate, y-axis and y-coordinate).	1
M.5.24	Represent real world and mathematical problems by graphing points in the first quadrant of the coordinate plane and interpret coordinate values of points in the context of the situation.	2
MD.G.M.5E	Classify two-dimensional figures into categories based on their properties.	
M.5.25	Understand that attributes belonging to a category of two-dimensional figures also belong to all subcategories of that category. For example, all rectangles have four right angles and squares are rectangles, so all squares have four right angles.	2
M.5.26	Classify two-dimensional figures in a hierarchy based on properties.	1

	West Virginia College and Career Ready Standards for Grade 6 Mathematics	
Standard	Description	DOK
RP.NS.M.6	Ratios and Proportional Relationships (RP) and The Number System (NS)	
RP.NS.M.6A	Understand ratio concepts and use ratio reasoning to solve problems.	
M.6.1	Understand the concept of a ratio and use ratio language to describe a ratio relationship between two quantities. (e.g., <i>"The ratio of wings to beaks in the bird house at the zoo was 2:1, because for every 2 wings there was 1 beak."</i> <i>"For every vote candidate A received, candidate C received nearly three votes."</i>)	1
M.6.2	Understand the concept of a unit rate a/b associated with a ratio $a:b$ with $b \neq 0$, and use rate language in the context of a ratio relationship. For example, <i>"This recipe has a ratio of 3 cups of flour to 4 cups of sugar, so there is $3/4$ cup of flour for each cup of sugar."</i> <i>"We paid \$75 for 15 hamburgers, which is a rate of \$5 per hamburger."</i> Expectations for unit rates in this grade are limited to non-complex fractions.	1
M.6.3	Use ratio and rate reasoning to solve real-world and mathematical problems, e.g., by reasoning about tables of equivalent ratios, tape diagrams, double number line diagrams, or equations. a. Make tables of equivalent ratios relating quantities with whole-number measurements, find missing values in the tables, and plot the pairs of values on the coordinate plane. Use tables to compare ratios. b. Solve unit rate problems including those involving unit pricing and constant speed. (e.g., <i>if it took 7 hours to mow 4 lawns, then at that rate, how many lawns could be mowed in 35 hours? At what rate were lawns being mowed?</i>) c. Find a percent of a quantity as a rate per 100 (e.g., 30% of a quantity means 30/100 times the quantity); solve problems involving finding the whole, given a part and the percent. d. Use ratio reasoning to convert measurement units; manipulate and transform units appropriately when multiplying or dividing quantities.	2

RP.NS.M.6B	Apply and extend previous understandings of multiplication and division to divide fractions by fractions.	
M.6.4	Interpret and compute quotients of fractions, and solve word problems involving division of fractions by fractions, e.g., by using visual fraction models and equations to represent the problem. <i>(e.g., create a story context for $(2/3) \div (3/4)$ and use a visual fraction model to show the quotient; use the relationship between multiplication and division to explain that $(2/3) \div (3/4) = 8/9$ because $3/4$ of $8/9$ is $2/3$. (In general, $(a/b) \div (c/d) = ad/bc$.)</i> How much chocolate will each person get if 3 people share $1/2$ lb of chocolate equally? How many $3/4$ -cup servings are in $2/3$ of a cup of yogurt? How wide is a rectangular strip of land with length $3/4$ mi and area $1/2$ square mi?)	2
RP.NS.M.6C	Compute fluently with multi-digit numbers and find common factors and multiples.	
M.6.5	Fluently divide multi-digit numbers using the standard algorithm.	1
M.6.6	Fluently add, subtract, multiply, and divide multi-digit decimals using the standard algorithm for each operation.	1
M.6.7	Find the greatest common factor of two whole numbers less than or equal to 100 and the least common multiple of two whole numbers less than or equal to 12. Use the distributive property to express a sum of two whole numbers 1–100 with a common factor as a multiple of a sum of two whole numbers with no common factor. <i>For example, express $36 + 8$ as $4(9 + 2)$.</i>	1

RP.NS.M.6D	Apply and extend previous understandings of numbers to the system of rational numbers.	
M.6.8	Understand that positive and negative numbers are used together to describe quantities having opposite directions or values (e.g., temperature above/below zero, elevation above/below sea level, credits/debits, positive/negative electric charge); use positive and negative numbers to represent quantities in real-world contexts, explaining the meaning of 0 in each situation.	2
M.6.9	Understand a rational number as a point on the number line. Extend number line diagrams and coordinate axes familiar from previous grades to represent points on the line and in the plane with negative number coordinates. a. Recognize opposite signs of numbers as indicating locations on opposite sides of 0 on the number line; recognize that the opposite of the opposite of a number is the number itself, e.g., $-(-3) = 3$, and that 0 is its own opposite. b. Understand signs of numbers in ordered pairs as indicating locations in quadrants of the coordinate plane; recognize that when two ordered pairs differ only by signs, the locations of the points are related by reflections across one or both axes. c. Find and position integers and other rational numbers on a horizontal or vertical number line diagram; find and position pairs of integers and other rational numbers on a coordinate plane.	1
M.6.10	Understand ordering and absolute value of rational numbers. a. Interpret statements of inequality as statements about the relative position of two numbers on a number line diagram. <i>For example, interpret $-3 > -7$ as a statement that -3 is located to the right of -7 on a number line oriented from left to right.</i> b. Write, interpret, and explain statements of order for rational numbers in real-world contexts. <i>For example, write $-3^{\circ}\text{C} > -7^{\circ}\text{C}$ to express the fact that -3°C is warmer than -7°C.</i> c. Understand the absolute value of a rational number as its distance from 0 on the number line; interpret absolute value as magnitude for a positive or negative quantity in a real-world situation. <i>For example, for an account balance of -30 dollars, write $-30 = 30$ to describe the size of the debt in dollars.</i> d. Distinguish comparisons of absolute value from statements about order. <i>For example, recognize that an account balance less than -30 dollars represents a debt greater than 30 dollars.</i>	2
M.6.11	Solve real-world and mathematical problems by graphing points in all four quadrants of the coordinate plane. Include use of coordinates and absolute value to find distances between points with the same first coordinate or the same second coordinate.	2

EE.M.6	Expressions and Equations	
EE.M.6A	Apply and extend previous understandings of arithmetic to algebraic expressions.	
M.6.12	Write and evaluate numerical expressions involving whole-number exponents.	1
M.6.13	Write, read, and evaluate expressions in which letters stand for numbers. a. Write expressions that record operations with numbers and with letters standing for numbers. <i>For example, express the calculation “Subtract y from 5” as $5 - y$.</i> b. Identify parts of an expression using mathematical terms (sum, term, product, factor, quotient, coefficient); view one or more parts of an expression as a single entity. <i>For example, describe the expression $2(8 + 7)$ as a product of two factors; view $(8 + 7)$ as both a single entity and a sum of two terms.</i> c. Evaluate expressions at specific values of their variables. Include expressions that arise from formulas used in real-world problems. Perform arithmetic operations, including those involving whole-number exponents, in the conventional order when there are no parentheses to specify a particular order: Order of Operations. <i>For example, use the formulas $V = s^3$ and $A = 6s^2$ to find the volume and surface area of a cube with sides of length $s = 1/2$.</i>	1
M.6.14	Apply the properties of operations to generate equivalent expressions. <i>For example, apply the distributive property to the expression $3(2 + x)$ to produce the equivalent expression $6 + 3x$; apply the distributive property to the expression $24x + 18y$ to produce the equivalent expression $6(4x + 3y)$; apply properties of operations to $y + y + y$ to produce the equivalent expression $3y$.</i>	1
M.6.15	Identify when two expressions are equivalent (i.e., when the two expressions name the same number regardless of which value is substituted into them). <i>For example, the expressions $y + y + y$ and $3y$ are equivalent because they name the same number regardless of which number y stands for.</i>	1
EE.M.6B	Reason about and solve one-variable equations and inequalities.	
M.6.16	Understand solving an equation or inequality as a process of answering a question: which values from a specified set, if any, make the equation or inequality true? Use substitution to determine whether a given number in a specified set makes an equation or inequality true.	1
M.6.17	Use variables to represent numbers and write expressions when solving a real-world or mathematical problem; understand that a variable can represent an unknown number, or, depending on the purpose at hand, any number in a specified set.	2

M.6.18	Solve real-world and mathematical problems by writing and solving equations of the form $x + p = q$ and $px = q$ for cases in which p , q and x are all nonnegative rational numbers.	2
M.6.19	Write an inequality of the form $x > c$ or $x < c$ to represent a constraint or condition in a real-world or mathematical problem. Recognize that inequalities of the form $x > c$ or $x < c$ have infinitely many solutions; represent solutions of such inequalities on number line diagrams.	2
EE.M.6C	Represent and analyze quantitative relationships between dependent and independent variables.	
M.6.20	Use variables to represent two quantities in a real-world problem that change in relationship to one another; write an equation to express one quantity, thought of as the dependent variable, in terms of the other quantity, thought of as the independent variable. Analyze the relationship between the dependent and independent variables using graphs and tables, and relate these to the equation. <i>For example, in a problem involving motion at constant speed, list and graph ordered pairs of distances and times, and write the equation $d = 65t$ to represent the relationship between distance and time.</i>	2
G.SP.M.6	Geometry (G) and Statistics and Probability (SP)	
G.SP.M.6.A	Solve real-world and mathematical problems involving area, surface area, and volume.	
M.6.21	Find the area of right triangles, other triangles, special quadrilaterals and polygons by composing into rectangles or decomposing into triangles and other shapes; apply these techniques in the context of solving real-world and mathematical problems.	2
M.6.22	Find the volume of a right rectangular prism with fractional edge lengths by packing it with unit cubes of the appropriate unit fraction edge lengths, and show that the volume is the same as would be found by multiplying the edge lengths of the prism. Apply the formulas $V = lwh$ and $V = bh$ to find volumes of right rectangular prisms with fractional edge lengths in the context of solving real-world and mathematical problems.	2
M.6.23	Draw polygons in the coordinate plane given coordinates for the vertices; use coordinates to find the length of a side joining points with the same first coordinate or the same second coordinate. Apply these techniques in the context of solving real-world and mathematical problems.	2
M.6.24	Represent three-dimensional figures using nets made up of rectangles and triangles, and use the nets to find the surface area of these figures. Apply these techniques in the context of solving real-world and mathematical problems.	2

G.SP.M.6.B	Develop understanding of statistical variability.	
M.6.25	Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. <i>For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students’ ages.</i>	2
M.6.26	Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center (mean/median), spread (range), and overall shape.	1
M.6.27	Recognize that a measure of center for a numerical data set summarizes all of its values with a single number.	1
G.SP.M.6.C	Summarize and describe distributions.	
M.6.28	Display numerical data in plots on a number line, including dot plots, histograms, and box plots.	1
M.6.29	Summarize numerical data sets in relation to their context, such as by: a. Reporting the number of observations. b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement. c. Giving quantitative measures of center (median and/or mean), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered. d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.	2

	West Virginia College and Career Ready Standards for Grade 7 Mathematics	
Standard	Description	DOK
RP.NS.M.7	Ratios and Proportional Relationships (RP) and The Number System (NS)	
RP.NS.M.7A	Analyze proportional relationships and use them to solve real-world and mathematical problems.	
M.7.1	Compute unit rates associated with ratios of fractions, including ratios of lengths, areas and other quantities measured in like or different units. <i>(e.g., if a person walks $\frac{1}{2}$ mile in each $\frac{1}{4}$ hour, compute the unit rate as the complex fraction $\frac{1/2}{1/4}$ miles per hour, equivalently 2 miles per hour.</i>	2
M.7.2	Recognize and represent proportional relationships between quantities. a. Decide whether two quantities are in a proportional relationship, e.g., by testing for equivalent ratios in a table or graphing on a coordinate plane and observing whether the graph is a straight line through the origin. b. Identify the constant of proportionality (unit rate) in tables, graphs, equations, diagrams, and verbal descriptions of proportional relationships. c. Represent proportional relationships by equations. <i>For example, if total cost t is proportional to the number n of items purchased at a constant price p, the relationship between the total cost and the number of items can be expressed as $t = pn$.</i> d. Explain what a point (x, y) on the graph of a proportional relationship means in terms of the situation, with special attention to the points $(0, 0)$ and $(1, r)$ where r is the unit rate.	2
M.7.3	Use proportional relationships to solve multistep ratio and percent problems. <i>Examples: simple interest, tax, markups and markdowns, gratuities and commissions, fees, percent increase and decrease, percent error.</i>	2

RP.NS.M.7B	Apply and extend previous understandings of operations with fractions to add, subtract, multiply, and divide rational numbers.	
M.7.4	Apply and extend previous understandings of addition and subtraction to add and subtract rational numbers; represent addition and subtraction on a horizontal or vertical number line diagram. - Describe situations in which opposite quantities combine to make 0. <i>For example, a hydrogen atom has 0 charge because its two constituents are oppositely charged.</i> - Understand $p + q$ as the number located a distance q from p, in the positive or negative direction depending on whether q is positive or negative. Show that a number and its opposite have a sum of 0 (are additive inverses). Interpret sums of rational numbers by describing real-world contexts. - Understand subtraction of rational numbers as adding the additive inverse, $p - q = p + (-q)$. Show that the distance between two rational numbers on the number line is the absolute value of their difference, and apply this principle in real-world contexts. - Apply properties of operations as strategies to add and subtract rational numbers.	2
M.7.5	Apply and extend previous understandings of multiplication and division and of fractions to multiply and divide rational numbers. - Understand that multiplication is extended from fractions to rational numbers by requiring that operations continue to satisfy the properties of operations, particularly the distributive property, leading to products such as $(-1)(-1) = 1$ and the rules for multiplying signed numbers. Interpret products of rational numbers by describing real-world contexts. - Understand that integers can be divided, provided that the divisor is not zero, and every quotient of integers (with non-zero divisor) is a rational number. If p and q are integers, then $-(p/q) = (-p)/q = p/(-q)$. Interpret quotients of rational numbers by describing real-world contexts. - Apply properties of operations as strategies to multiply and divide rational numbers. - Convert a rational number to a decimal using long division; know that the decimal form of a rational number terminates in 0s or eventually repeats.	2
M.7.6	Solve real-world and mathematical problems involving the four operations with rational numbers. Instructional note: computations with rational numbers extend the rules for manipulating fractions to complex fractions.	2

EE.M.7	Expressions and Equations	
EE.M.7A	Use properties of operations to generate equivalent expressions.	
M.7.7	Apply properties of operations as strategies to add, subtract, factor, and expand linear expressions with rational coefficients.	1
M.7.8	Understand that rewriting an expression in different forms in a problem context can shed light on the problem and how the quantities in it are related. <i>For example, $a + 0.05a = 1.05a$ means that “increase by 5%” is the same as “multiply by 1.05.”</i>	2
EE.M.7B	Solve real-life and mathematical problems using numerical and algebraic expressions and equations.	
M.7.9	Solve multi-step real-life and mathematical problems posed with positive and negative rational numbers in any form (whole numbers, fractions, and decimals), using tools strategically. Apply properties of operations to calculate with numbers in any form; convert between forms as appropriate; and assess the reasonableness of answers using mental computation and estimation strategies. <i>For example: If a woman making \$25 an hour gets a 10% raise, she will make an additional $\frac{1}{10}$ of her salary an hour, or \$2.50, for a new salary of \$27.50. If you want to place a towel bar $9\frac{3}{4}$ inches long in the center of a door that is $27\frac{1}{2}$ inches wide, you will need to place the bar about 9 inches from each edge; this estimate can be used as a check on the exact computation.</i>	2
M.7.10	Use variables to represent quantities in a real-world or mathematical problem, and construct simple equations and inequalities to solve problems by reasoning about the quantities. a. Solve word problems leading to equations of the form $px + q = r$ and $p(x + q) = r$, where p, q, and r are specific rational numbers. Solve equations of these forms fluently. Compare an algebraic solution to an arithmetic solution, identifying the sequence of the operations used in each approach. <i>For example, the perimeter of a rectangle is 54 cm. Its length is 6 cm. What is its width? An arithmetic solution similar to “$54 - 6 - 6$ divided by 2” may be compared with the reasoning involved in solving the equation $2w - 12 = 54$. An arithmetic solution similar to “$54/2 - 6$” may be compared with the reasoning involved in solving the equation $2(w - 6) = 54$.</i> b. Solve word problems leading to inequalities of the form $px + q > r$ or $px + q < r$, where p, q, and r are specific rational numbers. Graph the solution set of the inequality and interpret it in the context of the problem. <i>For example: As a salesperson, you are paid \$50 per week plus \$3 per sale. This week you want your pay to be at least \$100. Write an inequality for the number of sales you need to make, and describe the solutions.</i>	2

G.M.7	Geometry	
G.M.7.A	Draw, construct, and describe geometrical figures and describe the relationships between them.	
M.7.11	Solve problems involving scale drawings of geometric figures, including computing actual lengths and areas from a scale drawing and reproducing a scale drawing at a different scale.	2
M.7.12	Draw (freehand, with ruler and protractor, and with technology) geometric shapes with given conditions. Focus on constructing triangles from three measures of angles or sides, noticing when the conditions determine a unique triangle, more than one triangle, or no triangle.	2
M.7.13	Describe the two-dimensional figures that result from slicing three-dimensional figures, as in plane sections of right rectangular prisms and right rectangular pyramids.	2
G.M.7.B	Solve real-life and mathematical problems involving angle measure, area, surface area, and volume.	
M.7.14	Know the formulas for the area and circumference of a circle and use them to solve problems; give an informal derivation of the relationship between the circumference and area of a circle.	2
M.7.15	Use facts about supplementary, complementary, vertical, and adjacent angles in a multi-step problem to write and solve simple equations for an unknown angle in a figure.	2
M.7.16	Solve real-world and mathematical problems involving area, volume and surface area of two- and three-dimensional objects composed of triangles, quadrilaterals, polygons, cubes, and right prisms.	2
SP.M.7	Statistics and Probability	
SP.M.7.A	Use random sampling to draw inferences about a population.	
M.7.17	Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.	2
M.7.18	Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. <i>For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or prediction might be.</i>	3

SP.M.7.B	Draw informal comparative inferences about two populations.	
M.7.19	Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.	2
M.7.20	Summarize numerical data sets in relation to their context, such as by: 1. Reporting the number of observations. 2. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement. 3. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.	2
M.7.21	Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. (e.g., The mean height of players on the basketball team is 10 cm greater than the mean height of players on the soccer team, about twice the variability (mean absolute deviation) on either team; on a dot plot, the separation between the two distributions of heights is noticeable.)	2
M.7.22	Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. (e.g., Decide whether the words in a chapter of a seventh-grade science book are generally longer than the words in a chapter of a fourth-grade science book.)	2
SP.M.7.C	Investigate chance processes and develop, use, and, evaluate probability models.	
M.7.23	Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely and a probability near 1 indicates a likely event.	1

M.7.24	Approximate the probability of a chance event by collecting data on the chance process that produces it and observing its long-run relative frequency, and predict the approximate relative frequency given the probability. (e.g., When rolling a number cube 600 times, predict that a 3 or 6 would be rolled roughly 200 times, but probably not exactly 200 times.)	2
M.7.25	Develop a probability model and use it to find probabilities of events. Compare probabilities from a model to observed frequencies; if the agreement is not good, explain possible sources of the discrepancy. 1. Develop a uniform probability model by assigning equal probability to all outcomes, and use the model to determine probabilities of events. (e.g., If a student is selected at random from a class, find the probability that Jane will be selected and the probability that a girl will be selected.) 2. Develop a probability model (which may not be uniform) by observing frequencies in data generated from a chance process. (e.g., Find the approximate probability that a spinning penny will land heads up or that a tossed paper cup will land open-end down. Do the outcomes for the spinning penny appear to be equally likely based on the observed frequencies?)	3
M.7.26	Find probabilities of compound events using organized lists, tables, tree diagrams, and simulation. 1. Understand that, just as with simple events, the probability of a compound event is the fraction of outcomes in the sample space for which the compound event occurs. 2. Represent sample spaces for compound events using methods such as organized lists, tables and tree diagrams. For an event described in everyday language (e.g., “rolling double sixes”), identify the outcomes in the sample space which compose the event. 3. Design and use a simulation to generate frequencies for compound events. (e.g., Use random digits as a simulation tool to approximate the answer to the question: If 40% of donors have type A blood, what is the probability that it will take at least 4 donors to find one with type A blood?)	3

	West Virginia College and Career Ready Standards for Grade 8 Mathematics	
Standard	Description	DOK
NS.EE.M.8	The Number System (NS) and Expressions and Equations (EE)	
NS.EE.M.8.A	Know that there are numbers that are not rational, and approximate them by rational numbers.	
M.8.1	Know that numbers that are not rational are called irrational. Understand informally that every number has a decimal expansion; for rational numbers show that the decimal expansion repeats eventually, and convert a decimal expansion which repeats eventually into a rational number. Instructional Note: A decimal expansion that repeats the digit 0 is often referred to as a “terminating decimal.”	1
M.8.2	Use rational approximations of irrational numbers to compare the size of irrational numbers, locate them approximately on a number line diagram, and estimate the value of expressions (e.g., $\sqrt{2}$). <i>For example, by truncating the decimal expansion of $\sqrt{2}$, show that $\sqrt{2}$ is between 1 and 2, then between 1.4 and 1.5, and explain how to continue on to get better approximations.</i>	2
NS.EE.M.8.B	Work with radicals and integer exponents.	
M.8.3	Know and apply the properties of integer exponents to generate equivalent numerical expressions. <i>For example, $3^2 \cdot 3^{-5} = 3^{-3} = 1/3^3 = 1/27$.</i>	1
M.8.4	Use square root and cube root symbols to represent solutions to equations of the form $x^2 = p$ and $x^3 = p$, where p is a positive rational number. Evaluate square roots of small perfect squares and cube roots of small perfect cubes. Know that $\sqrt{2}$ is irrational.	1
M.8.5	Use numbers expressed in the form of a single digit times an integer power of 10 to estimate very large or very small quantities, and to express how many times as much one is than the other. <i>For example, estimate the population of the United States as 3×10^8 and the population of the world as 7×10^9, and determine that the world population is more than 20 times larger.</i>	2
M.8.6	Perform operations with numbers expressed in scientific notation, including problems where both decimal and scientific notation are used. Use scientific notation and choose units of appropriate size for measurements of very large or very small quantities (e.g., use millimeters per year for seafloor spreading). Interpret scientific notation that has been generated by technology.	1

NS.EE.M.8.C	Understand the connections between proportional relationships, lines, and linear equations.	
M.8.7	Graph proportional relationships, interpreting the unit rate as the slope of the graph. Compare two different proportional relationships represented in different ways. <i>For example, compare a distance-time graph to a distance-time equation to determine which of two moving objects has greater speed.</i>	2
M.8.8	Use similar triangles to explain why the slope m is the same between any two distinct points on a non-vertical line in the coordinate plane; derive the equation $y = mx$ for a line through the origin and the equation $y = mx + b$ for a line intercepting the vertical axis at b .	2
NS.EE.M.8.D	Analyze and solve linear equations and pairs of simultaneous linear equations.	
M.8.9	Solve linear equations in one variable. a. Give examples of linear equations in one variable with one solution, infinitely many solutions, or no solutions. Show which of these possibilities is the case by successively transforming the given equation into simpler forms, until an equivalent equation of the form $x = a$, $a = a$, or $a = b$ results (where a and b are different numbers). b. Solve linear equations and inequalities with rational number coefficients, including those whose solutions require expanding expressions using the distributive property and collecting like terms.	2
M.8.10	Analyze and solve pairs of simultaneous linear equations. a. Understand that solutions to a system of two linear equations in two variables correspond to points of intersection of their graphs, because points of intersection satisfy both equations simultaneously. b. Solve systems of two linear equations in two variables algebraically, and estimate solutions by graphing the equations. Solve simple cases by inspection. <i>For example, $3x + 2y = 5$ and $3x + 2y = 6$ have no solution because $3x + 2y$ cannot simultaneously be 5 and 6.</i> c. Solve real-world and mathematical problems leading to two linear equations in two variables. <i>For example, given coordinates for two pairs of points, determine whether the line through the first pair of points intersects the line through the second pair.</i>	2

F.M.8	Functions (F)	
F.M.8.A	Define, evaluate, and compare functions.	
M.8.11	Understand that a function is a rule that assigns to each input exactly one output. The graph of a function is the set of ordered pairs consisting of an input and the corresponding output.	1
M.8.12	Compare properties of two functions each represented in a different way (algebraically, graphically, numerically in tables, or by verbal descriptions). <i>For example, given a linear function represented by a table of values and a linear function represented by an algebraic expression, determine which function has the greater rate of change.</i>	2
M.8.13	Interpret the equation $y = mx + b$ as defining a linear function, whose graph is a straight line; give examples of functions that are not linear. <i>For example, the function $A = s^2$ giving the area of a square as a function of its side length is not linear because its graph contains the points (1,1), (2,4) and (3,9), which are not on a straight line.</i>	2
F.M.8.B	Use functions to model relationships between quantities.	
M.8.14	Construct a function to model a linear relationship between two quantities. Determine the rate of change and initial value of the function from a description of a relationship or from two (x, y) values, including reading these from a table or from a graph. Interpret the rate of change and initial value of a linear function in terms of the situation it models, and in terms of its graph or a table of values.	2
M.8.15	Describe qualitatively the functional relationship between two quantities by analyzing a graph (e.g., where the function is increasing or decreasing, linear or nonlinear). Sketch a graph that exhibits the qualitative features of a function that has been described verbally.	2
G.SP.M.8	Geometry (G) and Statistics and Probability (SP)	
G.SP.M.8.A	Understand congruence and similarity using physical models, transparencies, or geometry software.	
M.8.16	Verify experimentally the properties of rotations, reflections, and translations: a. Lines are taken to lines, and line segments to line segments of the same length. b. Angles are taken to angles of the same measure. c. Parallel lines are taken to parallel lines.	2
M.8.17	Understand that a two-dimensional figure is congruent to another if the second can be obtained from the first by a sequence of rotations, reflections, and translations; given two congruent figures, describe a sequence that exhibits the congruence between them.	2
M.8.18	Describe the effect of dilations, translations, rotations, and reflections on two-dimensional figures using coordinates.	2

M.8.19	Understand that a two-dimensional figure is similar to another if the second can be obtained from the first by a sequence of rotations, reflections, translations, and dilations; given two similar two-dimensional figures, describe a sequence that exhibits the similarity between them.	2
M.8.20	Use informal arguments to establish facts about the angle sum and exterior angle of triangles, about the angles created when parallel lines are cut by a transversal, and the angle-angle criterion for similarity of triangles. <i>For example, arrange three copies of the same triangle so that the sum of the three angles appears to form a line, and give an argument in terms of transversals why this is so.</i>	2
G.SP.M.8.B	Understand and apply the Pythagorean Theorem.	
M.8.21	Explain a proof of the Pythagorean Theorem and its converse.	2
M.8.22	Apply the Pythagorean Theorem to determine unknown side lengths in right triangles in real-world and mathematical problems in two and three dimensions.	2
M.8.23	Apply the Pythagorean Theorem to find the distance between two points in a coordinate system.	1
G.SP.M.8.C	Solve real-world and mathematical problems involving volume of cylinders, cones, and spheres.	
M.8.24	Know the formulas for the volumes of cones, cylinders, and spheres and use them to solve real-world and mathematical problems.	2
G.SP.M.8.D	Investigate patterns of association in bivariate data.	
M.8.25	Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association and nonlinear association.	2
M.8.26	Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.	2
M.8.27	Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. <i>For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.</i>	2
M.8.28	Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables. <i>For example, collect data from students in your class on whether or not they have a curfew on school nights and whether or not they have assigned chores at home. Is there evidence that those who have a curfew also tend to have chores?</i>	2

Appendix B

Data Analysis Tables

West Virginia College- and Career- Readiness Standards for Mathematics Grades 3-8 and Sample Test Events

2019

Brief Explanation of Data in the Alignment Tables by Column

Tables x.1

Domain # Number of Domains for each Reporting Category (RC)

Standards # Average number of standards for reviewers. If the number is greater than the actual number in the standard, then at least one reviewer coded an item for the Domain or RC but did not find any standard in the Domain that corresponded to the item.

Level The Depth-of-Knowledge levels coded by the reviewers for the standards for the RC.

Num of Stds by Level The number of standards coded at each DOK level
% w/in RC by Level The percent of standards within the RC coded at each DOK level

Hits

Mean & SD Mean and standard deviation number of items reviewers coded as corresponding to reporting category. The total is the total number of coded hits.

Categorical Concurrence:

“Yes” indicates that the reporting category met the acceptable level for criterion.

“Yes” if mean is six or more.

“Weak” if mean is five to six.

“No” if mean is less than five.

Tables x.2

First five columns repeat columns from Table 1.

DOK Level of Item Gives percentages of DOK levels of items in relation to standards

Mean percent and standard deviation of items coded as “under” the Depth-of-Knowledge level of the corresponding standard, as “at” (the same) the Depth-of-Knowledge level of the corresponding standard, and as “above” the Depth-of-Knowledge level of the corresponding standard.

DOK Consistency:

“Yes” indicates that 50% or more of the items were rated as “at” or “above” the Depth-of-Knowledge level of the corresponding standards.

“Weak” indicates that 40% to 50% of the items were rated as “at” or “above” the Depth-of-Knowledge level of the corresponding standards.

“No” indicates that less than 40% items were rated as “at” or “above” the Depth-of-Knowledge level of the corresponding standards.

Tables x.3

First five columns repeat columns from Table 1 and 2.

Range of Standards

Num Stds Hit Average number and standard deviation of the standards hit coded by reviewers.

% of Total Average percent and standard deviation of the total standards that had at least one item coded.

Range of Know(ledge):

“Yes” indicates that 50% or more of the standards had at least one coded standard.

“Weak” indicates that 40% to 50% of the standards had at least one coded standard.

“No” indicates that 40% or less of the standards had at least one coded standard.

Balance Index:

Gives % Hits in Std/Ttl Hit Average and standard deviation of the percent of the items hit for a reporting category of total number of hits (see total under the Hits column).

Index Average and standard deviation of the Balance Index.

Note: $BALANCE\ INDEX = 1 - (\sum_{k=1} |1/(O) - I_{(k)} / (H)|) / 2$

Where O = Total number of standards hit for the reporting category

$I_{(k)}$ = Number of items hit corresponding to

standard (k)

H = Total number of items hit for the reporting category

Balance of Representation:

“Yes” indicates that the Balance Index was .7 or above (items evenly distributed among standards).

“Weak” indicates that the Balance Index was .6 to .7 (a high percentage of items coded as corresponding to two or three standards).

“No” indicates that the Balance Index was .6 or less (a high percentage of items coded as corresponding to one standard.)

Tables x.4

Summary of if reporting category met the acceptable level for the four criteria by each standard.

Tables x.5

The DOK value for each assessment item given by each reviewer. The intraclass correlation for the group of reviewers is given on the last row.

Tables x.6

The DOK level and standard code assigned by each reviewer for each item.

Tables x.7

This lists for each standard all of the items coded by the group of reviewers as corresponding to the standard. The number of reviewers who coded the item is given in parentheses.

Tables x.8

This list for each item all of the standards coded by the group of reviewers as corresponding to the item. The number of reviewers who coded the standard is given in after the colon.

Tables x.9

This table can be used to compare approximately the DOK level of a standard to the average DOK level of the items reviewers assigned to the standard. This table is helpful to identify items with a lower DOK level that should be replaced by an item with a higher DOK level to improve the Depth-of-Knowledge Consistency. The DOK listed in the table for each item is generally the mode DOK for that item.

Batch 1 West Virginia Math Grade 3

Table 3.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B1_V2*
Number of Assessment Items - 34

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9.83	1 2	2 7	22.22 77.78	11.17	0.75	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	1 2	4 2	66.67 33.33	18.67	1.51	YES
MD.G.M.3 Measurement and Data and Geometry	5	10.17	1 2	2 8	20 80	9.33	1.03	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	19	34	1 2 3	8 19 6	24 58 18	39.17	0.41	

Table 3.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B1_V2*
Number of Assessment Items - 34

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9.83	11.17	0.75	49.9	21	45.68	22	4.42	8	WEAK
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	18.67	1.51	21.27	8	59.23	7	19.49	5	YES
MD.G.M.3 Measurement and Data and Geometry	5	10.17	9.33	1.03	42.99	17	38.7	18	18.31	9	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	19	34	39.17	0.41	34.89	12.2	50.21	11.4	14.89	4.2	

Table 3.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B1_V2 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9.83	11.17	0.75	7.5	0.55	76.3	4.95	YES	32	2	0.76	0.05	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	18.67	1.51	5.83	0.41	97.22	6.8	YES	45	4	0.84	0.05	YES
MD.G.M.3 Measurement and Data and Geometry	5	10.17	9.33	1.03	7	0.63	68.94	6.84	YES	24	3	0.84	0.03	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	19	34	39.17	0.41	5.1	3.46	60.61	42		25	19	0.81	0.1	

Table 3.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 3 B1_V2 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.3 Operations and Algebraic Thinking (OA)	YES	WEAK	YES	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.3 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 3.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 3 B1_V2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	1	2	1	1	1	1
2	1	1	1	1	1	1
3	1	1	1	1	1	1
4	1	1	1	1	1	1
5	1	1	1	1	1	1
6	1	2	1	1	1	1
7	2	2	3	2	2	2
8	2	2	2	2	2	2
9	1	2	1	1	1	1
10	2	2	3	2	2	2
11	1	2	2	1	1	1
12	1	2	1	1	1	1
13	1	1	1	1	1	1
14	1	1	1	1	1	1
15	1	2	2	2	2	2
16	1	2	1	1	1	1
17	1	1	1	1	1	1
18	1	2	1	1	1	2
19	1	1	1	1	1	1
20	1	1	1	1	1	1
21	1	2	1	1	1	1
22	1	2	1	1	1	1
23	3	3	3	3	3	2
24	1	2	1	1	1	1
25	1	2	1	1	1	1
26	1	1	1	1	1	1
27	2	2	2	2	1	2
28	1	1	1	1	1	1
29	1	2	1	1	2	1
30	1	1	1	1	1	1
31	1	2	1	1	1	1
32	1	2	1	1	1	1
33	2	2	1	2	2	2
34	1	1	1	1	1	1

Intraclass correlation - .9435

Pairwise Comparison - 0.79

Table 3.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 3 B1_V2 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	M.3.3			2	M.3.12			1	M.3.3			1	M.3.3			1	M.3.3			1	M.3.3		
2	1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13		
3	1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11		
4	1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10		
5	1	M.3.15			1	M.3.15			1	M.3.15			1	M.3.15			1	M.3.15			1	M.3.15		
6	1	M.3.14			2	M.3.14			1	M.3.14			1	M.3.14			1	M.3.19			1	M.3.14		
7	2	M.3.18			2	M.3.18			3	M.3.3			2	M.3.18			2	M.3.18			2	M.3.18		
8	2	M.3.3			2	M.3.12			2	M.3.12			2	M.3.12			2	M.3.3			2	M.3.12		
9	1	M.3.3			2	M.3.6			1	M.3.3			1	M.3.3			1	M.3.3			1	M.3.3		
10	2	M.3.8			2	M.3.3			3	M.3.4			2	M.3.8			2	M.3.8			2	M.3.8		
11	1	M.3.13			2	M.3.15			2	M.3.15			1	M.3.13			1	M.3.13			1	M.3.13		
12	1	M.3.17			2	M.3.17			1	M.3.17			1	M.3.17			1	M.3.17			1	M.3.17		
13	1	M.3.5			1	M.3.5			1	M.3.5			1	M.3.5			1	M.3.5			1	M.3.5		
14	1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11		
15	1	M.3.12			2	M.3.12			2	M.3.12			2	M.3.12			2	M.3.12			2	M.3.12		
16	1	M.3.23			2	M.3.23			1	M.3.23			1	M.3.23			1	M.3.20			1	M.3.23		
17	1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10		
18	1	M.3.15			2	M.3.15			1	M.3.15			1	M.3.15			1	M.3.13			2	M.3.15		
19	1	M.3.16			1	M.3.16			1	M.3.16			1	M.3.16			1	M.3.16			1	M.3.16		
20	1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13		
21	1	M.3.3			2	M.3.3			1	M.3.14			1	M.3.3			1	M.3.3			1	M.3.3		
22	1	M.3.5			2	M.3.8			1	M.3.4			1	M.3.5			1	M.3.4			1	M.3.5		
23	3	M.3.25	M.3.13		3	M.3.25	M.3.13		3	M.3.25	M.3.13		3	M.3.25	M.3.13		3	M.3.25	M.3.13		2	M.3.13	M.3.25	
24	1	M.3.18			2	M.3.18			1	M.3.18			1	M.3.18			1	M.3.18			1	MD.G.M.3B		
25	1	M.3.9			2	M.3.9			1	M.3.9			1	M.3.9			1	M.3.9			1	M.3.9		
26	1	M.3.7			1	M.3.7			1	M.3.7			1	M.3.7			1	M.3.1			1	M.3.7		
27	2	OA.M.3A			2	OA.M.3A			2	OA.M.3A			2	OA.M.3A			1	OA.M.3A			2	M.3.3		
28	1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13	M.3.25	
29	1	M.3.6			2	M.3.6			1	M.3.6			1	M.3.6			2	M.3.4			1	M.3.6		
30	1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11		
31	1	M.3.21			2	M.3.22			1	M.3.21			1	M.3.21			1	M.3.21			1	M.3.21		
32	1	M.3.14			2	M.3.14			1	M.3.14			1	M.3.14			1	M.3.19			1	M.3.14		
33	2	M.3.22			2	M.3.22			1	M.3.22			2	M.3.22			2	M.3.22			2	M.3.23		
34	1	M.3.4			1	M.3.4			1	M.3.4			1	M.3.4			1	M.3.4			1	M.3.4		
Objective Pairwise Comparison: 0.77																								
Standard Pairwise Comparison: 0.91																								

Table 3.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV MATH 2019 Grade 3 B1_V2

Low		Medium			High		
0		7.2			12		
OA.M.3							
OA.M.3A	27(5)						
M.3.1	26(1)						
M.3.2							
M.3.3	21(5)	9(5)	1(5)	7(1)	8(2)	10(1)	27(1)
M.3.4	34(6)	10(1)	22(2)	29(1)			
OA.M.3B							
M.3.5	22(3)	13(6)					
M.3.6	9(1)	29(5)					
OA.M.3C							
M.3.7	26(5)						
OA.M.3D							
M.3.8	22(1)	10(4)					
M.3.9	25(6)						
NBT.NF.M.3							
NBT.NF.M.3A							
M.3.10	17(6)	4(12)					
M.3.11	3(6)	14(6)	30(6)				
M.3.12	15(6)	8(4)	1(1)				
NBT.NF.M.3B							
M.3.13	2(6)	11(4)	18(1)	20(6)	28(6)	23(12)	

M.3.14	32(5)	21(1)	6(5)				
M.3.15	5(12)	11(2)	18(5)				
MD.G.M.3							
MD.G.M.3A							
M.3.16	19(6)						
M.3.17	12(6)						
MD.G.M.3B	24(1)						
M.3.18	24(5)	7(5)					
M.3.19	6(1)	32(1)					
MD.G.M.3C							
M.3.20	16(1)						
M.3.21	31(5)						
M.3.22	31(1)	33(5)					
MD.G.M.3D							
M.3.23	33(1)	16(5)					
MD.G.M.3E							
M.3.24							
M.3.25	23(12)	28(1)					
MHM							
MHM1							
MHM2							
MHM3							
MHM4							
MHM5							
MHM6							
MHM7							
MHM8							

Table 3.8 *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers) WV MATH 2019 Grade 3 B1_V2*

Low 2.4		Medium 7.2	High 12
1 REP10199-621	M.3.3:5	M.3.12:1	
2 10199-420	M.3.13:6		
3 REP10199-327	M.3.11:6		
4 10199-11551	M.3.10:12		
5 10199-7612	M.3.15:12		
6 10199-9968	M.3.14:5	M.3.19:1	
7 10199-342	M.3.3:1	M.3.18:5	
8 REP10199-10032	M.3.3:2	M.3.12:4	
9 10199-196	M.3.3:5	M.3.6:1	
10 10199-168	M.3.3:1	M.3.4:1	M.3.8:4
11 10199-934	M.3.13:4	M.3.15:2	
12 10199-334	M.3.17:6		
13 REP10199-9810	M.3.5:6		
14 10199-349	M.3.11:6		
15 10199-536	M.3.12:6		
16 10199-176	M.3.20:1	M.3.23:5	
17 10199-75	M.3.10:6		
18 10199-679	M.3.13:1	M.3.15:5	
19 10199-419	M.3.16:6		
20 10199-8362	M.3.13:6		
21 10199-10111	M.3.3:5	M.3.14:1	
22 10199-255	M.3.4:2	M.3.5:3	M.3.8:1
23 10199-9780	M.3.13:12	M.3.25:12	
24 10199-336	MD.G.M.3B:1	M.3.18:5	
25 10199-612	M.3.9:6		
26 10199-9717	M.3.1:1	M.3.7:5	
27 REP10199-267	OA.M.3A:5	M.3.3:1	
28 10199-130	M.3.13:6	M.3.25:1	
29 10199-277	M.3.4:1	M.3.6:5	
30 10199-100	M.3.11:6		
31 10199-346	M.3.21:5	M.3.22:1	
32 REP10199-941	M.3.14:5	M.3.19:1	
33 REP10199-263	M.3.22:5	M.3.23:1	
34 10199-214	M.3.4:6		

Table 3.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
 WV MATH 2019 Grade 3 B1_V2

Low DOK		Matched DOK		High DOK

OA.M.3							
OA.M.3A: [2]	27:(5)[2]						
M.3.1: [2]	26:(1)[1]						
M.3.2							
M.3.3: [2]	1:(5)[1]	7:(1)[3]	8:(2)[2]	9:(5)[1]	10:(1)[2]	21:(5)[1]	27:(1)[2]
M.3.4: [1]	10:(1)[3]	22:(2)[1]	29:(1)[2]	34:(6)[1]			
OA.M.3B							
M.3.5: [2]	13:(6)[1]	22:(3)[1]					
M.3.6: [2]	9:(1)[2]	29:(5)[1]					
OA.M.3C							
M.3.7: [1]	26:(5)[1]						
OA.M.3D							
M.3.8: [2]	10:(4)[2]	22:(1)[2]					
M.3.9: [2]	25:(6)[1]						
NBT.NF.M.3							
NBT.NF.M.3A							
M.3.10: [1]	4:(12)[1]	17:(6)[1]					
M.3.11: [1]	3:(6)[1]	14:(6)[1]	30:(6)[1]				
M.3.12: [1]	1:(1)[2]	8:(4)[2]	15:(6)[2]				
NBT.NF.M.3B							
M.3.13: [1]	2:(6)[1]	11:(4)[1]	18:(1)[1]	20:(6)[1]	23:(12)[3]	28:(6)[1]	
M.3.14: [2]	6:(5)[1]	21:(1)[1]	32:(5)[1]				
M.3.15: [2]	5:(12)[1]	11:(2)[2]	18:(5)[1]				
MD.G.M.3							
MD.G.M.3A							
M.3.16: [2]	19:(6)[1]						
M.3.17: [2]	12:(6)[1]						
MD.G.M.3B: [2]	24:(1)[1]						

M.3.18: [2]	7:(5)[2]	24:(5)[1]					
M.3.19: [2]	6:(1)[1]	32:(1)[1]					
MD.G.M.3C							
M.3.20: [1]	16:(1)[1]						
M.3.21: [1]	31:(5)[1]						
M.3.22: [2]	31:(1)[2]	33:(5)[2]					
MD.G.M.3D							
M.3.23: [2]	16:(5)[1]	33:(1)[2]					
MD.G.M.3E							
M.3.24							
M.3.25: [2]	23:(12)[3]	28:(1)[1]					
MHM							
MHM1							
MHM2							
MHM3							
MHM4							
MHM5							
MHM6							
MHM7							
MHM8							

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 3B1 MATH

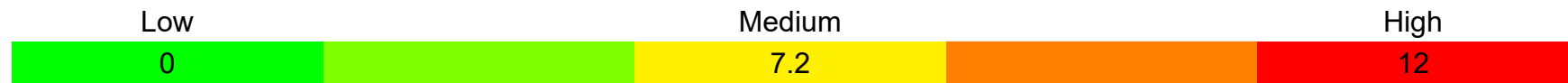
[illegible]

Table 3.CONF *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)*
CONF WV Grade 3B1 MATH

Low		Medium		High	
2.4		7.2		12	
1 REP10199-621	Exact:5	Partial:1			
2 10199-420	Exact:6				
3 REP10199-327	Exact:6				
4 10199-11551	Exact:12				
5 10199-7612	Exact:12				
6 10199-9968	Exact:5	Partial:1			
7 10199-342	Exact:4	Partial:1	Minimal:1		
8 REP10199-10032	Exact:5	Partial:1			
9 10199-196	Exact:6				
10 10199-168	Exact:6				
11 10199-934	Exact:6				
12 10199-334	Exact:6				
13 REP10199-9810	Exact:6				
14 10199-349	Exact:6				
15 10199-536	Exact:2	Partial:4			
16 10199-176	Exact:3	Partial:3			
17 10199-75	Exact:5	Minimal:1			
18 10199-679	Exact:4	Partial:2			
19 10199-419	Exact:6				
20 10199-8362	Exact:6				
21 10199-10111	Exact:6				
22 10199-255	Exact:5	Partial:1			
23 10199-9780	Exact:4	Partial:4	Minimal:4		
24 10199-336	Exact:3	Partial:2	Minimal:1		
25 10199-612	Exact:5	Partial:1			
26 10199-9717	Exact:6				
27 REP10199-267	Partial:4	Minimal:2			
28 10199-130	Exact:4	Partial:2			
29 10199-277	Exact:6				
30 10199-100	Exact:6				
31 10199-346	Exact:4	Partial:2			
32 REP10199-941	Exact:6				
33 REP10199-263	Exact:5	Partial:1			
34 10199-214	Exact:6				

Batch 2 West Virginia Math Grade 3

Table 3.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B2 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9	1 2	2 7	22.22 77.78	12.67	1.75	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	1 2	4 2	66.67 33.33	13.67	2.07	YES
MD.G.M.3 Measurement and Data and Geometry	5	10	1 2	2 8	20 80	10.33	0.82	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	19	33	1 2 3	8 19 6	24 58 18	36.67	0.82	

Table 3.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B2 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9	12.67	1.75	42.48	19	51.61	17	5.9	9	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	13.67	2.07	20.13	12	73.96	9	5.91	6	YES
MD.G.M.3 Measurement and Data and Geometry	5	10	10.33	0.82	40.56	26	59.44	26	0	0	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	19	33	36.67	0.82	33.64	16.1	61.82	14.7	4.55	3.4	

Table 3.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B2 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9	12.67	1.75	6.17	1.17	68.52	12.99	YES	33	4	0.75	0.03	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	13.67	2.07	5.83	0.41	97.22	6.8	YES	39	5	0.91	0.06	YES
MD.G.M.3 Measurement and Data and Geometry	5	10	10.33	0.82	6.33	1.51	63.33	15.06	YES	27	3	0.83	0.02	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	19	33	36.67	0.82	4.6	3.06	57.27	41		25	17	0.83	0.11	

Table 3.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 3 B2 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.3 Operations and Algebraic Thinking (OA)	YES	YES	YES	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.3 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 3.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 3 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	1	2	1	1	2	1
2	1	1	1	1	1	1
3	1	1	1	1	1	1
4	1	1	1	1	2	1
5	1	1	1	1	1	1
6	1	2	1	1	1	1
7	2	2	3	2	2	1
8	1	1	1	1	1	1
9	2	2	2	1	1	2
10	2	2	2	1	2	2
11	1	2	1	1	1	1
12	1	2	1	1	1	1
13	2	2	1	2	1	2
14	1	1	1	1	1	1
15	2	2	2	1	3	2
16	1	2	1	1	1	2
17	2	2	1	1	2	2
18	1	2	1	1	1	1
19	2	2	2	2	2	1
20	1	1	2	1	1	1
21	2	2	2	2	1	1
22	1	2	1	1	1	1
23	2	2	2	2	1	2
24	1	2	1	1	1	1
25	1	2	1	1	1	1
26	1	2	2	1	1	1
27	1	2	2	1	2	1
28	1	1	1	1	1	1
29	1	2	1	1	2	1
30	2	2	2	1	2	2
31	2	2	1	1	2	1
32	2	2	1	2	2	2
33	1	2	1	1	1	1
34	1	2	2	1	2	1

Intraclass correlation - .8032

Pairwise Comparison - 0.64

Table 3.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 3 B2 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	M.3.15			2	M.3.15			1	M.3.15			1	M.3.15			2	M.3.15			1	M.3.15		
2	1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11		
3	1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13		
4	1	M.3.12			1	M.3.3			1	M.3.3			1	M.3.3			2	M.3.3			1	M.3.3	M.3.12	
5	1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10		
6	1	M.3.14			2	M.3.14			1	M.3.14			1	M.3.14			1	M.3.19			1	M.3.14		
7	2	M.3.3			2	M.3.3			3	M.3.3			2	M.3.3			2	M.3.3			1	M.3.3	M.3.1	
8	1	M.3.12			1	M.3.12			1	M.3.12			1	M.3.12			1	M.3.7			1	M.3.12		
9	2	M.3.18			2	M.3.18			2	M.3.18			1	M.3.18			1	M.3.18			2	M.3.18		
10	2	M.3.8			2	M.3.3			2	M.3.8			1	M.3.8			2	M.3.8			2	M.3.3		
11	1	M.3.13			2	M.3.25			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13		
12	1	M.3.17			2	M.3.17			1	M.3.17			1	M.3.17			1	M.3.18			1	M.3.17		
13	2	M.3.5			2	M.3.5			1	M.3.5			2	M.3.5			1	M.3.5			2	M.3.5		
14	1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11		
15	2	M.3.16			2	M.3.16			2	M.3.16			1	M.3.16			3	M.3.8			2	M.3.16		
16	1	M.3.12			2	M.3.12			1	M.3.12			1	M.3.12			1	M.3.3			2	M.3.12		
17	2	M.3.9			2	M.3.9			1	M.3.9			1	M.3.9			2	M.3.9			2	M.3.9		
18	1	M.3.23			2	M.3.23			1	M.3.23			1	M.3.23			1	M.3.23			1	M.3.23		
19	2	M.3.25			2	M.3.22			2	M.3.13			2	M.3.25			2	M.3.22			1	M.3.13		
20	1	M.3.5			1	M.3.5			2	M.3.4			1	M.3.5			1	M.3.5			1	M.3.5		
21	2	M.3.15			2	M.3.15			2	M.3.15			2	M.3.15			1	M.3.15			1	M.3.15		
22	1	M.3.3			2	M.3.3			1	M.3.4			1	M.3.3			1	M.3.3			1	M.3.2		
23	2	M.3.18			2	M.3.18			2	M.3.18			2	M.3.18			1	M.3.18			2	M.3.18		
24	1	M.3.4			2	M.3.9			1	M.3.4			1	M.3.4			1	M.3.4			1	M.3.7		
25	1	M.3.19			2	M.3.19			1	M.3.19			1	M.3.19			1	M.3.19			1	M.3.19		
26	1	M.3.3			2	M.3.3			2	M.3.1			1	M.3.3			1	M.3.3			1	M.3.3		
27	1	M.3.6			2	M.3.6			2	M.3.6			1	M.3.6			2	M.3.4			1	M.3.6		
28	1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10		
29	1	M.3.3			2	M.3.3			1	M.3.6			1	M.3.3			2	M.3.6			1	M.3.3		
30	2	M.3.23			2	M.3.23			2	M.3.23			1	M.3.23			2	M.3.22			2	M.3.23		
31	2	M.3.24	M.3.13		2	M.3.13			1	M.3.25			1	M.3.25	M.3.13		2	M.3.22			1	M.3.13		
32	2	M.3.22			2	M.3.22			1	M.3.22			2	M.3.22			2	M.3.22			2	M.3.23		
33	1	M.3.14			2	M.3.14			1	M.3.13			1	M.3.14			1	M.3.14			1	M.3.14		
34	1	M.3.15			2	M.3.15			2	M.3.15			1	M.3.15			2	M.3.14			1	M.3.15		
Objective Pairwise Comparison: 0.74																								
Standard Pairwise Comparison: 0.91																								

Table 3.7

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)

WV MATH 2019 Grade 3 B2

Low		Medium				High	
0		6				10	
OA.M.3							
OA.M.3A							
M.3.1	26(1)	7(1)					
M.3.2	22(1)						
M.3.3	22(4)	16(1)	26(5)	29(4)	4(5)	7(6)	10(2)
M.3.4	24(8)	27(1)	22(1)	20(1)			
OA.M.3B							
M.3.5	20(5)	13(6)					
M.3.6	27(5)	29(2)					
OA.M.3C							
M.3.7	24(2)	8(1)					
OA.M.3D							
M.3.8	15(2)	10(4)					
M.3.9	24(2)	17(6)					
NBT.NF.M.3							
NBT.NF.M.3A							
M.3.10	28(6)	5(6)					
M.3.11	14(6)	2(6)					
M.3.12	8(5)	16(5)	4(2)				
NBT.NF.M.3B							
M.3.13	19(2)	3(6)	11(5)	33(1)	31(4)		

M.3.14	33(5)	34(1)	6(5)				
M.3.15	1(6)	21(6)	34(5)				
MD.G.M.3							
MD.G.M.3A							
M.3.16	15(10)						
M.3.17	12(5)						
MD.G.M.3B							
M.3.18	12(1)	9(6)	23(6)				
M.3.19	25(6)	6(1)					
MD.G.M.3C							
M.3.20							
M.3.21							
M.3.22	30(1)	19(2)	31(1)	32(5)			
MD.G.M.3D							
M.3.23	32(1)	18(6)	30(5)				
MD.G.M.3E							
M.3.24	31(1)						
M.3.25	31(2)	19(2)	11(1)				
MHM							
MHM1							
MHM2							
MHM3							
MHM4							
MHM5							
MHM6							
MHM7							
MHM8							

Table 3.8 *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers) WV MATH 2019 Grade 3 B2*

Low 2		Medium 6		High 10	
1 10199-823	M.3.15:6				
2 10199-286	M.3.11:6				
3 10199-527	M.3.13:6				
4 REP10199-621	M.3.3:5	M.3.12:2			
5 10199-58	M.3.10:6				
6 REP10199-941	M.3.14:5	M.3.19:1			
7 REP10199-257	M.3.1:1	M.3.3:6			
8 10199-156	M.3.7:1	M.3.12:5			
9 10199-338	M.3.18:6				
10 10199-433	M.3.3:2	M.3.8:4			
11 10199-515	M.3.13:5	M.3.25:1			
12 10199-322	M.3.17:5	M.3.18:1			
13 10199-253	M.3.5:6				
14 10199-359	M.3.11:6				
15 REP10199-237	M.3.8:2	M.3.16:10			
16 10199-622	M.3.3:1	M.3.12:5			
17 10199-514	M.3.9:6				
18 10199-11570	M.3.23:6				
19 10199-413	M.3.13:2	M.3.22:2	M.3.25:2		
20 REP10199-9810	M.3.4:1	M.3.5:5			
21 10199-504	M.3.15:6				
22 REP10199-546	M.3.2:1	M.3.3:4	M.3.4:1		
23 10199-467	M.3.18:6				
24 REP10199-256	M.3.4:8	M.3.7:2	M.3.9:2		
25 10199-185	M.3.19:6				
26 10199-227	M.3.1:1	M.3.3:5			
27 10199-308	M.3.4:1	M.3.6:5			
28 REP10199-66	M.3.10:6				
29 REP10199-580	M.3.3:4	M.3.6:2			
30 10199-170	M.3.22:1	M.3.23:5			
31 10199-333	M.3.13:4	M.3.22:1	M.3.24:1	M.3.25:2	
32 REP10199-263	M.3.22:5	M.3.23:1			
33 10199-454	M.3.13:1	M.3.14:5			
34 10199-9882	M.3.14:1	M.3.15:5			

Table 3.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV MATH 2019 Grade 3 B2

Low DOK		Matched DOK		High DOK

OA.M.3							
OA.M.3A							
M.3.1: [2]	7:(1)[1]	26:(1)[2]					
M.3.2: [2]	22:(1)[1]						
M.3.3: [2]	4:(5)[1]	7:(6)[2]	10:(2)[2]	16:(1)[1]	22:(4)[1]	26:(5)[1]	29:(4)[1]
M.3.4: [1]	20:(1)[2]	22:(1)[1]	24:(8)[1]	27:(1)[2]			
OA.M.3B							
M.3.5: [2]	13:(6)[2]	20:(5)[1]					
M.3.6: [2]	27:(5)[1]	29:(2)[2]					
OA.M.3C							
M.3.7: [1]	8:(1)[1]	24:(2)[1]					
OA.M.3D							
M.3.8: [2]	10:(4)[2]	15:(2)[3]					
M.3.9: [2]	17:(6)[2]	24:(2)[2]					
NBT.NF.M.3							
NBT.NF.M.3A							
M.3.10: [1]	5:(6)[1]	28:(6)[1]					
M.3.11: [1]	2:(6)[1]	14:(6)[1]					
M.3.12: [1]	4:(2)[1]	8:(5)[1]	16:(5)[1]				
NBT.NF.M.3B							
M.3.13: [1]	3:(6)[1]	11:(5)[1]	19:(2)[2]	31:(4)[2]	33:(1)[1]		
M.3.14: [2]	6:(5)[1]	33:(5)[1]	34:(1)[2]				
M.3.15: [2]	1:(6)[1]	21:(6)[2]	34:(5)[1]				
MD.G.M.3							
MD.G.M.3A							
M.3.16: [2]	15:(10)[2]						
M.3.17: [2]	12:(5)[1]						
MD.G.M.3B							

M.3.18: [2]	9:(6)[2]	12:(1)[1]	23:(6)[2]				
M.3.19: [2]	6:(1)[1]	25:(6)[1]					
MD.G.M.3C							
M.3.20							
M.3.21							
M.3.22: [2]	19:(2)[2]	30:(1)[2]	31:(1)[2]	32:(5)[2]			
MD.G.M.3D							
M.3.23: [2]	18:(6)[1]	30:(5)[2]	32:(1)[2]				
MD.G.M.3E							
M.3.24: [2]	31:(1)[2]						
M.3.25: [2]	11:(1)[2]	19:(2)[2]	31:(2)[1]				
MHM							
MHM1							
MHM2							
MHM3							
MHM4							
MHM5							
MHM6							
MHM7							
MHM8							

Batch 3 West Virginia Math Grade 3

Table 3.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B3 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9.5	1 2	2 7	22.22 77.78	12.67	1.86	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	1 2	4 2	66.67 33.33	15.83	1.83	YES
MD.G.M.3 Measurement and Data and Geometry	5	10	1 2	2 8	20 80	9.17	0.98	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	19	33.5	1 2 3	8 19 6	24 58 18	37.67	1.63	

Table 3.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B3 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9.5	12.67	1.86	23.66	13	76.34	13	0	0	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	15.83	1.83	17.9	7	65.6	10	16.5	7	YES
MD.G.M.3 Measurement and Data and Geometry	5	10	9.17	0.98	30.81	20	69.19	20	0	0	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	19	33.5	37.67	1.63	23.01	10.5	69.91	12.1	7.08	3.1	

Table 3.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B3 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9.5	12.67	1.86	6.67	0.82	70.19	8.12	YES	34	5	0.7	0.05	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	15.83	1.83	5.83	0.41	97.22	6.8	YES	43	5	0.86	0.02	YES
MD.G.M.3 Measurement and Data and Geometry	5	10	9.17	0.98	6.67	0.82	66.67	8.16	YES	24	3	0.83	0.02	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	19	33.5	37.67	1.63	4.8	3.22	58.52	41		25	18	0.8	0.12	

Table 3.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 3 B3 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.3 Operations and Algebraic Thinking (OA)	YES	YES	YES	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.3 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 3.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 3 B3

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	2	1	1	1	1
2	1	1	1	1	1	1
3	2	2	2	2	2	2
4	1	1	1	1	1	1
5	1	1	1	1	1	1
6	2	2	2	3	2	2
7	2	2	2	2	1	2
8	1	2	1	1	1	1
9	2	2	2	2	2	2
10	2	2	1	2	2	1
11	2	1	2	1	2	2
12	2	2	2	1	2	2
13	1	1	2	1	1	1
14	1	1	1	1	1	1
15	1	2	1	1	1	2
16	2	2	2	2	2	2
17	1	2	2	1	1	1
18	1	1	1	1	1	1
19	2	2	1	1	2	2
20	1	1	1	1	1	1
21	2	2	2	2	2	2
22	1	2	1	1	1	1
23	2	2	2	2	2	1
24	2	2	1	1	1	1
25	2	2	2	1	2	2
26	1	2	1	1	2	1
27	2	1	2	2	2	2
28	2	2	2	2	2	2
29	2	2	2	2	2	2
30	1	2	1	1	1	1
31	1	1	1	1	2	1
32	2	2	1	2	2	2
33	1	2	1	1	2	1
34	2	2	2	1	2	2

Intraclass correlation - .8849

Pairwise Comparison - 0.73

Table 3.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 3 B3 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.3.15			2	M.3.15			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.15		
2	1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10		
3	2	M.3.3			2	M.3.12			2	M.3.12			2	M.3.12			2	M.3.3			2	M.3.12	M.3.3	
4	1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11		
5	1	M.3.15			1	M.3.13			1	M.3.15			1	M.3.15			1	M.3.15			1	M.3.15		
6	2	M.3.14			2	M.3.14			2	M.3.14			3	M.3.14			2	M.3.14			2	M.3.14		
7	2	M.3.18			2	M.3.18			2	M.3.18			2	M.3.18			1	M.3.18			2	M.3.18		
8	1	M.3.12			2	M.3.12			1	M.3.12			1	M.3.12			1	M.3.7			1	M.3.12		
9	2	M.3.8			2	M.3.3			2	M.3.8			2	M.3.8			2	M.3.3			2	M.3.3		
10	2	M.3.3			2	M.3.3			1	M.3.3			2	M.3.3			2	M.3.6			1	M.3.3	M.3.6	
11	2	M.3.25			1	M.3.13			2	M.3.13			1	M.3.13			2	M.3.15			2	M.3.13		
12	2	M.3.16			2	M.3.16			2	M.3.16			1	M.3.16			2	M.3.16			2	M.3.16		
13	1	M.3.5			1	M.3.5			2	M.3.5			1	M.3.5			1	M.3.5			1	M.3.5		
14	1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11		
15	1	M.3.12			2	M.3.3			1	M.3.12			1	M.3.12			1	M.3.3			2	M.3.12	M.3.3	
16	2	M.3.17			2	M.3.17			2	M.3.17			2	M.3.17			2	M.3.11			2	M.3.17		
17	1	M.3.23			2	M.3.23			2	M.3.23			1	M.3.23			1	M.3.23			1	M.3.23		
18	1	M.3.15			1	M.3.15			1	M.3.15			1	M.3.15			1	M.3.15			1	M.3.15		
19	2	M.3.9			2	M.3.9			1	M.3.9			1	M.3.9			2	M.3.9			2	M.3.9		
20	1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13		
21	2	M.3.24			2	M.3.24			2	M.3.24			2	M.3.24			2	M.3.24			2	M.3.24		
22	1	M.3.4			2	M.3.9			1	M.3.4			1	M.3.4			1	M.3.4			1	M.3.7		
23	2	M.3.3			2	M.3.3			2	M.3.3			2	M.3.3			2	M.3.3			1	M.3.3	M.3.1	
24	2	M.3.18			2	M.3.18			1	M.3.18			1	M.3.18			1	M.3.18			1	M.3.18		
25	2	M.3.6			2	M.3.6			2	M.3.6			1	M.3.6			2	M.3.6			2	M.3.6		
26	1	M.3.3			2	M.3.3			1	M.3.6			1	M.3.3			2	M.3.6			1	M.3.3		
27	2	M.3.15			1	M.3.13			2	M.3.15			2	M.3.15			2	M.3.15			2	M.3.15		
28	2	M.3.10			2	M.3.10			2	M.3.14			2	M.3.10			2	M.3.10			2	M.3.10		
29	2	OA.M.3A			2	M.3.3			2	OA.M.3A			2	OA.M.3A			2	M.3.3			2	M.3.3		
30	1	M.3.19			2	M.3.14			1	M.3.19			1	M.3.19			1	M.3.18			1	M.3.19		
31	1	M.3.4			1	M.3.4			1	M.3.4			1	M.3.4			2	M.3.6			1	M.3.4		
32	2	M.3.22			2	M.3.22			1	M.3.22			2	M.3.22			2	M.3.22			2	M.3.23		
33	1	M.3.14			2	M.3.14			1	M.3.14			1	M.3.14			2	M.3.19			1	M.3.14		
34	2	M.3.25			2	M.3.15			2	M.3.15			1	M.3.15			2	M.3.8			2	M.3.15		
Objective Pairwise Comparison: 0.73																								
Standard Pairwise Comparison: 0.9																								

Table 3.7

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)

WV MATH 2019 Grade 3 B3

Low		Medium			High		
0		7.2			12		
OA.M.3							
OA.M.3A	29(3)						
M.3.1	23(1)						
M.3.2							
M.3.3	29(3)	26(4)	23(6)	10(5)	9(3)	3(3)	15(3)
M.3.4	22(8)	31(5)					
OA.M.3B							
M.3.5	13(6)						
M.3.6	31(1)	25(6)	26(2)	10(2)			
OA.M.3C							
M.3.7	22(2)	8(1)					
OA.M.3D							
M.3.8	9(3)	34(1)					
M.3.9	22(2)	19(6)					
NBT.NF.M.3							
NBT.NF.M.3A							
M.3.10	2(6)	28(5)					
M.3.11	4(6)	14(6)	16(1)				
M.3.12	8(5)	3(4)	15(4)				
NBT.NF.M.3B							

M.3.13	20(6)	27(1)	5(1)	1(3)	11(4)		
M.3.14	6(12)	30(1)	28(1)	33(5)			
M.3.15	34(4)	27(5)	18(6)	1(3)	5(5)	11(1)	
MD.G.M.3							
MD.G.M.3A							
M.3.16	12(12)						
M.3.17	16(5)						
MD.G.M.3B							
M.3.18	7(6)	24(6)	30(1)				
M.3.19	30(4)	33(1)					
MD.G.M.3C							
M.3.20							
M.3.21							
M.3.22	32(5)						
MD.G.M.3D							
M.3.23	32(1)	17(6)					
MD.G.M.3E							
M.3.24	21(6)						
M.3.25	11(1)	34(1)					
MHM							
MHM1							
MHM2							
MHM3							
MHM4							
MHM5							
MHM6							
MHM7							
MHM8							

Table 3.8 *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers) WV MATH 2019 Grade 3 B3*

Low 2.4		Medium 7.2	High 12	
1 10199-750	M.3.13:3	M.3.15:3		
2 10199-205	M.3.10:6			
3 REP10199-10032	M.3.3:3	M.3.12:4		
4 10199-351	M.3.11:6			
5 10199-549	M.3.13:1	M.3.15:5		
6 10199-795	M.3.14:12			
7 10199-340	M.3.18:6			
8 10199-9880	M.3.7:1	M.3.12:5		
9 10199-9900	M.3.3:3	M.3.8:3		
10 10199-9957	M.3.3:5	M.3.6:2		
11 10199-508	M.3.13:4	M.3.15:1	M.3.25:1	
12 REP10199-237	M.3.16:12			
13 10199-9774	M.3.5:6			
14 10199-101	M.3.11:6			
15 10199-9896	M.3.3:3	M.3.12:4		
16 10199-10132	M.3.11:1	M.3.17:5		
17 10199-157	M.3.23:6			
18 10199-510	M.3.15:6			
19 10199-474	M.3.9:6			
20 10199-566	M.3.13:6			
21 REP10199-453	M.3.24:6			
22 REP10199-256	M.3.4:8	M.3.7:2	M.3.9:2	
23 REP10199-257	M.3.1:1	M.3.3:6		
24 10199-337	M.3.18:6			
25 10199-188	M.3.6:6			
26 REP10199-580	M.3.3:4	M.3.6:2		
27 10199-284	M.3.13:1	M.3.15:5		
28 10199-84	M.3.10:5	M.3.14:1		
29 REP10199-267	OA.M.3A:3	M.3.3:3		
30 10199-609	M.3.14:1	M.3.18:1	M.3.19:4	
31 REP10199-162	M.3.4:5	M.3.6:1		
32 REP10199-263	M.3.22:5	M.3.23:1		
33 10199-9869	M.3.14:5	M.3.19:1		
34 10199-11561	M.3.8:1	M.3.15:4	M.3.25:1	

Table 3.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV MATH 2019 Grade 3 B3

Low DOK		Matched DOK		High DOK

OA.M.3							
OA.M.3A: [2]	29:(3)[2]						
M.3.1: [2]	23:(1)[1]						
M.3.2							
M.3.3: [2]	3:(3)[2]	9:(3)[2]	10:(5)[2]	15:(3)[2]	23:(6)[2]	26:(4)[1]	29:(3)[2]
M.3.4: [1]	22:(8)[1]	31:(5)[1]					
OA.M.3B							
M.3.5: [2]	13:(6)[1]						
M.3.6: [2]	10:(2)[2]	25:(6)[2]	26:(2)[2]	31:(1)[2]			
OA.M.3C							
M.3.7: [1]	8:(1)[1]	22:(2)[1]					
OA.M.3D							
M.3.8: [2]	9:(3)[2]	34:(1)[2]					
M.3.9: [2]	19:(6)[2]	22:(2)[2]					
NBT.NF.M.3							
NBT.NF.M.3A							
M.3.10: [1]	2:(6)[1]	28:(5)[2]					
M.3.11: [1]	4:(6)[1]	14:(6)[1]	16:(1)[2]				
M.3.12: [1]	3:(4)[2]	8:(5)[1]	15:(4)[1]				
NBT.NF.M.3B							
M.3.13: [1]	1:(3)[1]	5:(1)[1]	11:(4)[2]	20:(6)[1]	27:(1)[1]		
M.3.14: [2]	6:(12)[2]	28:(1)[2]	30:(1)[2]	33:(5)[1]			
M.3.15: [2]	1:(3)[2]	5:(5)[1]	11:(1)[2]	18:(6)[1]	27:(5)[2]	34:(4)[2]	
MD.G.M.3							
MD.G.M.3A							
M.3.16: [2]	12:(12)[2]						
M.3.17: [2]	16:(5)[2]						

MD.G.M.3B							
M.3.18: [2]	7:(6)[2]	24:(6)[1]	30:(1)[1]				
M.3.19: [2]	30:(4)[1]	33:(1)[2]					
MD.G.M.3C							
M.3.20							
M.3.21							
M.3.22: [2]	32:(5)[2]						
MD.G.M.3D							
M.3.23: [2]	17:(6)[1]	32:(1)[2]					
MD.G.M.3E							
M.3.24: [2]	21:(6)[2]						
M.3.25: [2]	11:(1)[2]	34:(1)[2]					
MHM							
MHM1							
MHM2							
MHM3							
MHM4							
MHM5							
MHM6							
MHM7							
MHM8							

Table 3.CONF

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 3B3 MATH

Low										Medium										High											
0										4.8										8											
Agreement with internal coding																															
Exact	4(4)	5(4)	6(8)	7(3)	9(4)	10(4)	11(4)	12(8)	13(3)	14(4)	15(4)	16(3)	17(4)	18(4)	19(3)	20(4)	21(3)	8(3)	22(4)	23(4)	24(4)	25(3)	26(1)	27(2)	28(3)	29(3)	30(3)	31(3)	32(4)	33(3)	34(4)
Partial	33(1)	31(1)	29(1)	28(1)	26(2)	25(1)	22(4)	8(1)	16(1)	3(2)	13(1)	7(1)																			
Minimal	19(1)	26(1)	27(2)	30(1)																											
Negligible	21(1)																														

Table 3.CONF *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)*
CONF WV Grade 3B3 MATH

Low	Medium	High
1.6	4.8	8

1 10199-750	Exact:4	
2 10199-205	Exact:4	
3 REP10199-10032	Exact:2	Partial:2
4 10199-351	Exact:4	
5 10199-549	Exact:4	
6 10199-795	Exact:8	
7 10199-340	Exact:3	Partial:1
8 10199-9880	Exact:3	Partial:1
9 10199-9900	Exact:4	
10 10199-9957	Exact:4	
11 10199-508	Exact:4	
12 REP10199-237	Exact:8	
13 10199-9774	Exact:3	Partial:1
14 10199-101	Exact:4	
15 10199-9896	Exact:4	
16 10199-10132	Exact:3	Partial:1
17 10199-157	Exact:4	
18 10199-510	Exact:4	
19 10199-474	Exact:3	Minimal:1
20 10199-566	Exact:4	
21 REP10199-453	Exact:3	Negligible:1
22 REP10199-256	Exact:4	Partial:4
23 REP10199-257	Exact:4	
24 10199-337	Exact:4	
25 10199-188	Exact:3	Partial:1
26 REP10199-580	Exact:1	Partial:2
27 10199-284	Exact:2	Minimal:2
28 10199-84	Exact:3	Partial:1
29 REP10199-267	Exact:3	Partial:1
30 10199-609	Exact:3	Minimal:1
31 REP10199-162	Exact:3	Partial:1
32 REP10199-263	Exact:4	
33 10199-9869	Exact:3	Partial:1
34 10199-11561	Exact:4	

Batch 4 West Virginia Math Grade 3

Table 3.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B4 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9	1 2	2 7	22.22 77.78	10.33	1.03	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	1 2	4 2	66.67 33.33	15.33	2.5	YES
MD.G.M.3 Measurement and Data and Geometry	5	10	1 2	2 8	20 80	8.83	1.47	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	19	33	1 2 3	8 19 6	24 58 18	34.49	0.84	

Table 3.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B4 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9	10.33	1.03	25.29	15	74.71	15	0	0	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	15.33	2.5	9.35	7	82.7	10	7.96	7	YES
MD.G.M.3 Measurement and Data and Geometry	5	10	8.83	1.47	22.28	15	73.78	16	3.94	6	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	19	33	34.49	0.84	17.39	7.4	77.78	8	4.83	3.8	

Table 3.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 3 B4 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.3 Operations and Algebraic Thinking (OA)	4	9	10.33	1.03	7	1.1	77.78	12.17	YES	30	3	0.77	0.08	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	2	6	15.33	2.5	5.83	0.41	97.22	6.8	YES	44	7	0.87	0.06	YES
MD.G.M.3 Measurement and Data and Geometry	5	10	8.83	1.47	7.67	1.03	76.67	10.33	YES	26	4	0.89	0.06	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	19	33	34.49	0.84	5.1	3.5	62.92	43		25	18	0.84	0.09	

Table 3.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 3 B4 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.3 Operations and Algebraic Thinking (OA)	YES	YES	YES	YES
NBT.NF.M.3 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.3 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 3.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 3 B4

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	1	1	1	1	1	1
2	1	1	2	1	1	1
3	1	1	1	1	1	1
4	1	1	1	1	1	1
5	1	2	2	2	2	1
6	2	2	2	1	1	2
7	2	2	2	2	2	2
8	1	1	1	1	1	1
9	2	2	1	1	1	2
10	2	2	2	2	2	2
11	1	1	1	1	1	1
12	2	2	2	2	2	2
13	2	2	1	2	2	1
14	1	1	1	1	1	1
15	1	2	1	1	1	1
16	2	2	2	2	2	2
17	2	2	2	2	2	2
18	2	1	1	1	2	2
19	1	1	1	1	1	1
20	1	2	2	2	2	2
21	2	2	2	2	1	2
22	1	2	1	2	2	1
23	2	1	2	2	2	2
24	1	2	1	1	2	1
25	1	2	1	1	2	1
26	2	2	2	1	2	1
27	2	3	2	1	1	1
28	2	2	2	2	2	1
29	1	1	1	1	1	1
30	2	2	1	2	2	2
31	1	1	1	1	1	1
32	2	2	2	3	2	2
33	2	2	2	2	2	2
34	1	1	1	1	1	1

Intraclass correlation - .8902

Pairwise Comparison - 0.75

Table 3.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 3 B4 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	M.3.12			1	M.3.12			1	M.3.12			1	M.3.12			1	M.3.3			1	M.3.12		
2	1	M.3.13			1	M.3.13			2	M.3.15			1	M.3.13			1	M.3.15			1	M.3.13		
3	1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11		
4	1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10		
5	1	M.3.15			2	M.3.15			2	M.3.15			2	M.3.15			2	M.3.15			1	M.3.15		
6	2	M.3.14			2	M.3.14			2	M.3.14			1	M.3.14			1	M.3.19			2	M.3.13		
7	2	M.3.3			2	M.3.3			2	M.3.3			2	M.3.3			2	M.3.3			2	M.3.3		
8	1	M.3.12			1	M.3.12			1	M.3.12			1	M.3.12			1	M.3.12			1	M.3.12		
9	2	M.3.16			2	M.3.16			1	M.3.16			1	M.3.16			1	M.3.16			2	M.3.16		
10	2	M.3.8			2	M.3.3			2	M.3.8			2	M.3.8			2	M.3.3			2	M.3.11		
11	1	M.3.25			1	M.3.13			1	M.3.13			1	M.3.13	M.3.25		1	M.3.14			1	M.3.13		
12	2	M.3.18			2	M.3.18			2	M.3.18			2	M.3.18			2	M.3.18			2	M.3.18		
13	2	M.3.5			2	M.3.5			1	M.3.5			2	M.3.5			2	M.3.5			1	M.3.5		
14	1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10		
15	1	M.3.12			2	M.3.12			1	M.3.3			1	M.3.12			1	M.3.3			1	M.3.12	M.3.3	
16	2	M.3.23			2	M.3.23			2	M.3.23			2	M.3.23			2	M.3.23			2	M.3.23		
17	2	M.3.24			2	M.3.24			2	M.3.24			2	M.3.24			2	M.3.24			2	M.3.24		
18	2	M.3.13			1	M.3.13			1	M.3.13			1	M.3.13			2	M.3.15			2	M.3.13	M.3.25	
19	1	M.3.7			1	M.3.1			1	M.3.7			1	M.3.7			1	M.3.1			1	M.3.7		
20	1	M.3.17			2	M.3.17			2	M.3.18			2	M.3.17			2	M.3.17			2	M.3.17		
21	2	M.3.9			2	M.3.1			2	M.3.9			2	M.3.9			1	M.3.9			2	M.3.9		
22	1	M.3.15			2	M.3.15			1	M.3.15			2	M.3.15			2	M.3.15			1	M.3.15		
23	2	M.3.8			1	M.3.11			2	M.3.8			2	M.3.8			2	M.3.3			2	M.3.3		
24	1	M.3.3			2	M.3.3			1	M.3.3			1	M.3.3			2	M.3.3			1	M.3.2		
25	1	M.3.19			2	M.3.14			1	M.3.19			1	M.3.19			2	M.3.19			1	M.3.19		
26	2	M.3.6			2	M.3.6			2	M.3.3			1	M.3.6			2	M.3.6			1	M.3.6		
27	2	M.3.22			3	M.3.22			2	M.3.13			1	M.3.25			1	M.3.21			1	M.3.13		
28	2	M.3.2			2	M.3.3			2	M.3.2			2	M.3.2			2	M.3.2			1	M.3.2		
29	1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11			1	M.3.11		
30	2	M.3.25			2	M.3.25			1	M.3.13			2	M.3.13			2	M.3.25			2	M.3.13		
31	1	M.3.4			1	M.3.4			1	M.3.4			1	M.3.4			1	M.3.4			1	M.3.4		
32	2	M.3.22			2	M.3.22			2	M.3.22			3	M.3.22			2	M.3.22			2	M.3.22		
33	2	M.3.14			2	M.3.14			2	M.3.14			2	M.3.14			2	M.3.19			2	M.3.14		
34	1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10			1	M.3.10		
Objective Pairwise Comparison: 0.74																								
Standard Pairwise Comparison: 0.87																								

Table 3.7

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)

WV MATH 2019 Grade 3 B4

	Low		Medium		High			
	0		3.6				6	
OA.M.3								
OA.M.3A								
M.3.1	19(2)	21(1)						
M.3.2	24(1)	28(5)						
M.3.3	28(1)	24(5)	26(1)	1(1)	7(6)	10(2)	23(2)	15(3)
M.3.4	31(6)							
OA.M.3B								
M.3.5	13(6)							
M.3.6	26(5)							
OA.M.3C								
M.3.7	19(4)							
OA.M.3D								
M.3.8	10(3)	23(3)						
M.3.9	21(5)							
NBT.NF.M.3								
NBT.NF.M.3A								
M.3.10	14(6)	4(6)	34(6)					
M.3.11	23(1)	10(1)	3(6)	29(6)				
M.3.12	15(4)	1(5)	8(6)					
NBT.NF.M.3B								

M.3.13	2(4)	6(1)	11(4)	30(3)	27(2)	18(5)		
M.3.14	25(1)	11(1)	6(4)	33(5)				
M.3.15	2(2)	5(6)	18(1)	22(6)				
MD.G.M.3								
MD.G.M.3A								
M.3.16	9(6)							
M.3.17	20(5)							
MD.G.M.3B								
M.3.18	20(1)	12(6)						
M.3.19	25(5)	6(1)	33(1)					
MD.G.M.3C								
M.3.20								
M.3.21	27(1)							
M.3.22	27(2)	32(6)						
MD.G.M.3D								
M.3.23	16(6)							
MD.G.M.3E								
M.3.24	17(6)							
M.3.25	27(1)	30(3)	11(2)	18(1)				
MHM								
MHM1								
MHM2								
MHM3								
MHM4								
MHM5								
MHM6								
MHM7								
MHM8								

Table 3.8 *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers) WV MATH 2019 Grade 3 B4*

	Low 1.2	Medium 3.6	High 6
1 10199-147	M.3.3:1	M.3.12:5	
2 10199-949	M.3.13:4	M.3.15:2	
3 REP10199-327	M.3.11:6		
4 REP10199-66	M.3.10:6		
5 10199-659	M.3.15:6		
6 10199-928	M.3.13:1	M.3.14:4	M.3.19:1
7 10199-179	M.3.3:6		
8 10199-148	M.3.12:6		
9 10199-625	M.3.16:6		
10 10199-9879	M.3.3:2	M.3.8:3	M.3.11:1
11 10199-10001	M.3.13:4	M.3.14:1	M.3.25:2
12 10199-341	M.3.18:6		
13 10199-248	M.3.5:6		
14 10199-204	M.3.10:6		
15 10199-9813	M.3.3:3	M.3.12:4	
16 10199-422	M.3.23:6		
17 REP10199-453	M.3.24:6		
18 10199-10037	M.3.13:5	M.3.15:1	M.3.25:1
19 10199-10169	M.3.1:2	M.3.7:4	
20 10199-11565	M.3.17:5	M.3.18:1	
21 10199-490	M.3.1:1	M.3.9:5	
22 10199-593	M.3.15:6		
23 10199-9953	M.3.3:2	M.3.8:3	M.3.11:1
24 REP10199-546	M.3.2:1	M.3.3:5	
25 10199-506	M.3.14:1	M.3.19:5	
26 10199-292	M.3.3:1	M.3.6:5	
27 10199-9779	M.3.13:2	M.3.21:1	M.3.22:2 M.3.25:1
28 10199-529	M.3.2:5	M.3.3:1	
29 10199-350	M.3.11:6		
30 10199-414	M.3.13:3	M.3.25:3	
31 REP10199-162	M.3.4:6		
32 10199-304	M.3.22:6		
33 10199-799	M.3.14:5	M.3.19:1	
34 10199-180	M.3.10:6		

Table 3.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV MATH 2019 Grade 3 B4

Low DOK		Matched DOK		High DOK

OA.M.3								
OA.M.3A								
M.3.1: [2]	19:(2)[1]	21:(1)[2]						
M.3.2: [2]	24:(1)[1]	28:(5)[2]						
M.3.3: [2]	1:(1)[1]	7:(6)[2]	10:(2)[2]	15:(3)[1]	23:(2)[2]	24:(5)[1]	26:(1)[2]	28:(1)[2]
M.3.4: [1]	31:(6)[1]							
OA.M.3B								
M.3.5: [2]	13:(6)[2]							
M.3.6: [2]	26:(5)[2]							
OA.M.3C								
M.3.7: [1]	19:(4)[1]							
OA.M.3D								
M.3.8: [2]	10:(3)[2]	23:(3)[2]						
M.3.9: [2]	21:(5)[2]							
NBT.NF.M.3								
NBT.NF.M.3A								
M.3.10: [1]	4:(6)[1]	14:(6)[1]	34:(6)[1]					
M.3.11: [1]	3:(6)[1]	10:(1)[2]	23:(1)[1]	29:(6)[1]				
M.3.12: [1]	1:(5)[1]	8:(6)[1]	15:(4)[1]					
NBT.NF.M.3B								
M.3.13: [1]	2:(4)[1]	6:(1)[2]	11:(4)[1]	18:(5)[1]	27:(2)[2]	30:(3)[2]		
M.3.14: [2]	6:(4)[2]	11:(1)[1]	25:(1)[2]	33:(5)[2]				
M.3.15: [2]	2:(2)[2]	5:(6)[2]	18:(1)[2]	22:(6)[2]				
MD.G.M.3								
MD.G.M.3A								
M.3.16: [2]	9:(6)[2]							
M.3.17: [2]	20:(5)[2]							
MD.G.M.3B								
M.3.18: [2]	12:(6)[2]	20:(1)[2]						

M.3.19: [2]	6:(1)[1]	25:(5)[1]	33:(1)[2]					
MD.G.M.3C								
M.3.20								
M.3.21: [1]	27:(1)[1]							
M.3.22: [2]	27:(2)[2]	32:(6)[2]						
MD.G.M.3D								
M.3.23: [2]	16:(6)[2]							
MD.G.M.3E								
M.3.24: [2]	17:(6)[2]							
M.3.25: [2]	11:(2)[1]	18:(1)[2]	27:(1)[1]	30:(3)[2]				
MHM								
MHM1								
MHM2								
MHM3								
MHM4								
MHM5								
MHM6								
MHM7								
MHM8								

Batch 1 West Virginia Math Grade 4

Table 4.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 4 B1 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.4 Operations and Algebraic Thinking (OA)	3	5	1 2	1 4	20 80	12	2.28	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	5	13	1 2	6 7	46.15 53.85	15.33	2.34	YES
MD.G.M.4 Measurement and Data and Geometry	4	10	1 2	6 4	60 40	9	0.89	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	1.17	0.98	NO
Total	20	36	1 2 3	13 17 6	36 47 17	37.5	1.97	

Table 4.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 4 B1 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.4 Operations and Algebraic Thinking (OA)	3	5	12	2.28	47.26	15	49.96	13	2.78	4	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	5	13	15.33	2.34	30.36	7	55.39	10	14.26	8	YES
MD.G.M.4 Measurement and Data and Geometry	4	10	9	0.89	25.46	13	67.04	11	7.5	10	YES
MHM Mathematical Habits of Mind	8	8	1.17	0.98	13.33	30	86.67	30	0	0	YES
Total	20	36	37.5	1.97	34.67	7.8	56.44	6.2	8.89	5.7	

Table 4.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 4 B1 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.4 Operations and Algebraic Thinking (OA)	3	5	12	2.28	4.83	0.41	96.67	8.16	YES	31	7	0.87	0.04	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	5	13	15.33	2.34	12.17	1.17	93.59	8.99	YES	41	5	0.85	0.04	YES
MD.G.M.4 Measurement and Data and Geometry	4	10	9	0.89	8	1.26	80	12.65	YES	25	2	0.89	0.06	YES
MHM Mathematical Habits of Mind	8	8	1.17	0.98	1.17	0.98	14.58	12.29	NO	3	2	0.96	0.1	YES
Total	20	36	37.5	1.97	6.5	4.68	71.21	38		25	16	0.89	0.04	

Table 4.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 4 B1 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.4 Operations and Algebraic Thinking (OA)	YES	YES	YES	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.4 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	YES	NO	YES

Table 4.5 *Depth-of-Knowledge Levels by Item and Reviewers*
Intraclass Correlation WV MATH 2019 Grade 4 B1

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	1	2	1	1	1	1
2	1	1	1	1	1	1
3	3	3	3	3	2	3
4	1	1	1	1	1	1
5	1	1	1	1	1	1
6	1	2	1	1	2	1
7	2	2	1	2	2	2
8	2	2	1	2	1	2
9	2	2	2	2	1	2
10	1	1	1	1	1	1
11	1	1	1	1	1	1
12	1	1	1	1	2	2
13	2	2	2	1	2	2
14	1	1	1	1	1	1
15	1	1	1	1	1	1
16	1	2	2	1	1	2
17	1	2	1	2	2	1
18	2	1	1	1	1	1
19	2	1	1	1	2	2
20	1	1	1	1	1	1
21	1	1	1	1	1	1
22	1	1	1	1	1	1
23	2	2	2	2	2	2
24	1	2	1	1	1	1
25	1	1	1	1	1	1
26	1	1	1	1	1	1
27	1	1	1	1	1	1
28	1	2	1	1	1	1
29	2	2	1	1	1	2
30	1	2	1	1	2	2
31	1	1	1	1	1	1
32	1	2	1	1	2	2
33	1	2	1	1	1	1
34	1	1	1	1	1	1

Intraclass correlation - .9087

Pairwise Comparison - 0.76

Table 4.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 4 B1 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	M.4.1			2	M.4.1			1	M.4.1			1	M.4.1			1	M.4.1			1	M.4.1		
2	1	M.4.2			1	M.4.2			1	M.4.2			1	M.4.2			1	M.4.2			1	M.4.2		
3	3	M.4.15	M.4.18	MHM3	3	M.4.15	M.4.18	MHM3	3	M.4.15	M.4.18	MHM5	3	M.4.18	M.4.15	MHM3	2	M.4.1			3	M.4.15	M.4.18	MHM3
4	1	M.4.21			1	M.4.21			1	M.4.21			1	M.4.21			1	M.4.21			1	M.4.21		
5	1	M.4.12			1	M.4.12			1	M.4.12			1	M.4.12			1	M.4.12			1	M.4.12		
6	1	M.4.22			2	M.4.22			1	M.4.22			1	M.4.22			2	M.4.22			1	M.4.22		
7	2	M.4.11			2	M.4.3			1	M.4.3			2	M.4.3			2	M.4.3			2	M.4.3		
8	2	M.4.25			2	M.4.25			1	M.4.25			2	M.4.24	M.4.25		1	M.4.24			2	M.4.24	M.4.25	
9	2	M.4.3			2	M.4.2			2	M.4.2			2	M.4.3			1	M.4.3			2	M.4.3		
10	1	M.4.7			1	M.4.6			1	M.4.7			1	M.4.7			1	M.4.7			1	M.4.7		
11	1	M.4.10			1	M.4.10			1	M.4.10			1	M.4.10			1	M.4.10			1	M.4.10		
12	1	M. 4.20			1	M.4.9			1	M. 4.20			1	M. 4.20			2	M.4.2			2	M. 4.20		
13	2	M.4.13			2	M.4.13			2	M.4.13			1	M.4.13			2	M.4.12			2	M.4.13		
14	1	M.4.4			1	M.4.4			1	M.4.4			1	M.4.4			1	M.4.5			1	M.4.4		
15	1	M.4.26			1	M.4.26			1	M.4.26			1	M.4.26			1	M.4.26			1	M.4.26		
16	1	M.4.18			2	M.4.18			2	M.4.18			1	M.4.18			1	M.4.13			2	M.4.18		
17	1	M.4.2			2	M.4.1			1	M.4.1			2	M.4.2			2	M.4.2			1	M.4.2		
18	2	M.4.8			1	M.4.8			1	M.4.8			1	M.4.8			1	M.4.8			1	M.4.8		
19	2	M.4.14			1	M.4.14			1	M.4.14			1	M.4.14			2	M.4.3			2	M.4.14		
20	1	M.4.5			1	M.4.5			1	M.4.5			1	M.4.5			1	M.4.5			1	M.4.5		
21	1	M.4.11			1	M.4.11			1	M.4.11			1	M.4.11			1	M.4.2			1	M.4.11		
22	1	M.4.24			1	M.4.24			1	M.4.24			1	M.4.24			1	M.4.24			1	M.4.24		
23	2	M.4.3			2	M.4.3			2	M.4.2			2	M.4.3			2	M.4.2			2	M.4.3		
24	1	M.4.6			2	M.4.1			1	M.4.1			1	M.4.6			1	M.4.6			1	M.4.2	M.4.11	
25	1	M.4.17			1	M.4.17			1	M.4.17			1	M.4.17			1	M.4.17			1	M.4.17		
26	1	M.4.28			1	M.4.28			1	M.4.28			1	M.4.28			1	M.4.28			1	M.4.28		
27	1	M.4.16			1	M.4.16			1	M.4.16			1	M.4.16			1	M.4.16			1	M.4.16		
28	1	M.4.27			2	M.4.26			1	M.4.27			1	M.4.27			1	M.4.23			1	M.4.27		
29	2	M.4.9			2	M.4.9			1	M.4.9			1	M.4.9			1	M.4.9			2	M.4.9		
30	1	M.4.4			2	M.4.4			1	M.4.4			1	M.4.4			2	M.4.5			2	M.4.4		
31	1	M.4.7			1	M.4.7			1	M.4.7			1	M.4.7			1	M.4.3			1	M.4.7		
32	1	M.4.15			2	M.4.15			1	M.4.14			1	M.4.15			2	M.4.14			2	M.4.15	MHM2	
33	1	M.4.24			2	M.4.23			1	M.4.23			1	M.4.23			1	M.4.23			1	M.4.23		
34	1	M.4.5			1	M.4.5			1	M.4.5			1	M.4.5			1	M.4.5			1	M.4.5	MHM4	
Objective Pairwise Comparison: 0.73																								
Standard Pairwise Comparison: 0.9																								

Table 4.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers) WV MATH 2019 Grade 4 B1

	Low		Medium		High	
	0		6			10
OA.M.4						
OA.M.4A						
M.4.1	1(6)	3(1)	17(2)	24(2)		
M.4.2	24(1)	17(4)	21(1)	23(2)	2(6)	9(2)
M.4.3	9(4)	7(10)	23(4)	19(1)	31(1)	
OA.M.4B						
M.4.4	30(5)	14(5)				
OA.M.4C						
M.4.5	14(1)	30(1)	20(6)	34(6)		
NBT.NF.M.4						
NBT.NF.M.4A						
M.4.6	24(3)	10(1)				
M.4.7	10(5)	31(5)				
M.4.8	18(6)					
NBT.NF.M.4B						
M.4.9	29(6)	12(1)				
M.4.10	11(6)					
M.4.11	7(2)	21(5)	24(1)			
NBT.NF.M.4C						
M.4.12	5(6)	13(1)				
M.4.13	13(5)	16(1)				
NBT.NF.M.4D						
M.4.14	19(5)	32(2)				
M.4.15	32(4)	3(5)				

NBT.NF.M.4E							
M.4.16	27(6)						
M.4.17	25(6)						
M.4.18	16(5)	3(5)					
MD.G.M.4							
MD.G.M.4A							
M.4.19							
M. 4.20	12(4)						
M.4.21	4(6)						
MD.G.M.4B							
M.4.22	6(6)						
MD.G.M.4C							
M.4.23	28(1)	33(5)					
M.4.24	33(1)	22(6)	8(3)				
M.4.25	8(5)						
MD.G.M.4D							
M.4.26	15(6)	28(1)					
M.4.27	28(4)						
M.4.28	26(6)						
MHM							
MHM1							
MHM2	32(1)						
MHM3	3(4)						
MHM4	34(1)						
MHM5	3(1)						
MHM6							
MHM7							
MHM8							

Table 4.8 Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV MATH 2019 Grade 4 B1

Low		Medium		High	
2		6		10	
1 10199-9794	M.4.1:6				
2 10199-651	M.4.2:6				
3 10199-905	M.4.1:1	M.4.15:5	M.4.18:5	MHM3:4	MHM5:1
4 10199-779	M.4.21:6				
5 10199-9839	M.4.12:6				
6 REP10199-962	M.4.22:6				
7 REP10199-9728	M.4.3:10	M.4.11:2			
8 10199-873	M.4.24:3	M.4.25:5			
9 REP10199-678	M.4.2:2	M.4.3:4			
10 10199-386	M.4.6:1	M.4.7:5			
11 10199-598	M.4.10:6				
12 10199-2217	M.4.2:1	M.4.9:1	M. 4.20:4		
13 10199-850	M.4.12:1	M.4.13:5			
14 10199-450	M.4.4:5	M.4.5:1			
15 10199-614	M.4.26:6				
16 10199-1226	M.4.13:1	M.4.18:5			
17 10199-11470	M.4.1:2	M.4.2:4			
18 10199-589	M.4.8:6				
19 REP10199-9920	M.4.3:1	M.4.14:5			
20 10199-667	M.4.5:6				
21 10199-392	M.4.2:1	M.4.11:5			
22 10199-703	M.4.24:6				
23 10199-895	M.4.2:2	M.4.3:4			
24 10199-382	M.4.1:2	M.4.2:1	M.4.6:3	M.4.11:1	
25 10199-998	M.4.17:6				
26 10199-469	M.4.28:6				
27 10199-730	M.4.16:6				
28 10199-428	M.4.23:1	M.4.26:1	M.4.27:4		
29 10199-705	M.4.9:6				
30 REP10199-11526	M.4.4:5	M.4.5:1			
31 10199-317	M.4.3:1	M.4.7:5			
32 REP10199-9822	M.4.14:2	M.4.15:4	MHM2:1		
33 10199-846	M.4.23:5	M.4.24:1			
34 10199-944	M.4.5:6	MHM4:1			

Table 4.9 Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK]) WV MATH 2019 Grade 4 B1

Low DOK		Matched DOK		High DOK

OA.M.4							
OA.M.4A							
M.4.1: [2]	1:(6)[1]	3:(1)[2]	17:(2)[2]	24:(2)[2]			
M.4.2: [2]	2:(6)[1]	9:(2)[2]	12:(1)[2]	17:(4)[2]	21:(1)[1]	23:(2)[2]	24:(1)[1]
M.4.3: [2]	7:(10)[2]	9:(4)[2]	19:(1)[2]	23:(4)[2]	31:(1)[1]		
OA.M.4B							
M.4.4: [1]	14:(5)[1]	30:(5)[1]					
OA.M.4C							
M.4.5: [2]	14:(1)[1]	20:(6)[1]	30:(1)[2]	34:(6)[1]			
NBT.NF.M.4							
NBT.NF.M.4A							
M.4.6: [1]	10:(1)[1]	24:(3)[1]					
M.4.7: [1]	10:(5)[1]	31:(5)[1]					
M.4.8: [1]	18:(6)[1]						
NBT.NF.M.4B							
M.4.9: [1]	12:(1)[1]	29:(6)[2]					
M.4.10: [2]	11:(6)[1]						
M.4.11: [2]	7:(2)[2]	21:(5)[1]	24:(1)[1]				
NBT.NF.M.4C							
M.4.12: [2]	5:(6)[1]	13:(1)[2]					
M.4.13: [2]	13:(5)[2]	16:(1)[1]					
NBT.NF.M.4D							
M.4.14: [2]	19:(5)[1]	32:(2)[2]					
M.4.15: [2]	3:(5)[3]	32:(4)[2]					
NBT.NF.M.4E							
M.4.16: [1]	27:(6)[1]						
M.4.17: [1]	25:(6)[1]						
M.4.18: [2]	3:(5)[3]	16:(5)[2]					
MD.G.M.4							
MD.G.M.4A							
M.4.19							

M. 4.20: [2]	12:(4)[1]					
M.4.21: [2]	4:(6)[1]					
MD.G.M.4B						
M.4.22: [2]	6:(6)[1]					
MD.G.M.4C						
M.4.23: [1]	28:(1)[1]	33:(5)[1]				
M.4.24: [1]	8:(3)[2]	22:(6)[1]	33:(1)[1]			
M.4.25: [2]	8:(5)[2]					
MD.G.M.4D						
M.4.26: [1]	15:(6)[1]	28:(1)[2]				
M.4.27: [1]	28:(4)[1]					
M.4.28: [1]	26:(6)[1]					
MHM						
MHM1						
MHM2: [3]	32:(1)[2]					
MHM3: [3]	3:(4)[3]					
MHM4: [3]	34:(1)[1]					
MHM5: [3]	3:(1)[3]					
MHM6						
MHM7						
MHM8						

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 4B1 MATH

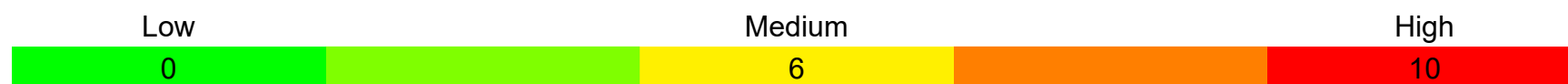
[illegible]

Table 4.CONF *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)*
CONF WV Grade 4B1 MATH

Low		Medium		High	
2		6		10	
1	10199-9794	Exact:6			
2	10199-651	Exact:6			
3	10199-905	Exact:1	Partial:4	Minimal:1	
4	10199-779	Exact:6			
5	10199-9839	Exact:3	Partial:3		
6	REP10199-962	Exact:5	Partial:1		
7	REP10199-9728	Exact:10	Partial:2		
8	10199-873	Exact:5	Partial:1		
9	REP10199-678	Exact:6			
10	10199-386	Exact:6			
11	10199-598	Exact:6			
12	10199-2217	Exact:5	Partial:1		
13	10199-850	Exact:5	Partial:1		
14	10199-450	Exact:5	Partial:1		
15	10199-614	Exact:6			
16	10199-1226	Exact:6			
17	10199-11470	Exact:6			
18	10199-589	Exact:6			
19	REP10199-9920	Exact:5	Partial:1		
20	10199-667	Exact:6			
21	10199-392	Exact:6			
22	10199-703	Exact:3	Partial:3		
23	10199-895	Exact:6			
24	10199-382	Exact:5	Negligible:1		
25	10199-998	Exact:5	Partial:1		
26	10199-469	Exact:6			
27	10199-730	Exact:6			
28	10199-428	Exact:3	Partial:3		
29	10199-705	Exact:4	Partial:2		
30	REP10199-11526	Exact:5	Partial:1		
31	10199-317	Exact:6			
32	REP10199-9822	Exact:5	Partial:1		
33	10199-846	Exact:3	Partial:2	Minimal:1	
34	10199-944	Exact:6			

Batch 2 West Virginia Math Grade 4

Table 4.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 4 B2 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.4 Operations and Algebraic Thinking (OA)	3	5	1 2	1 4	20 80	9.17	1.17	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	5	13.83	1 2	6 7	46.15 53.85	17.17	1.94	YES
MD.G.M.4 Measurement and Data and Geometry	4	10.17	1 2	6 4	60 40	9.17	1.94	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0.83	2.04	NO
Total	20	37	1 2 3	13 17 6	36 47 17	36.34	2.8	

Table 4.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 4 B2 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.4 Operations and Algebraic Thinking (OA)	3	5	9.17	1.17	40.89	25	51.24	21	7.87	12	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	5	13.83	17.17	1.94	26.74	16	61.75	19	11.51	6	YES
MD.G.M.4 Measurement and Data and Geometry	4	10.17	9.17	1.94	25.46	15	60.97	14	13.57	8	YES
MHM Mathematical Habits of Mind	8	8	0.83	2.04	100	NaN	0	NaN	0	NaN	NO
Total	20	37	36.34	2.8	31.65	15.6	57.34	14.9	11.01	5.3	

Table 4.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 4 B2 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.4 Operations and Algebraic Thinking (OA)	3	5	9.17	1.17	4.33	1.21	86.67	24.22	YES	26	4	0.85	0.09	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	5	13.83	17.17	1.94	11.5	1.22	83.06	7.62	YES	46	5	0.81	0.02	YES
MD.G.M.4 Measurement and Data and Geometry	4	10.17	9.17	1.94	8.17	0.75	80.3	6.37	YES	26	6	0.93	0.12	YES
MHM Mathematical Habits of Mind	8	8	0.83	2.04	0.5	1.22	6.25	15.31	NO	2	4	0.98	0.05	YES
Total	20	37	36.34	2.8	6.1	4.76	64.07	39		25	18	0.89	0.08	

Table 4.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 4 B2 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.4 Operations and Algebraic Thinking (OA)	YES	YES	YES	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.4 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NO	NO	YES

Table 4.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 4 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	1	2	1	1	2	1
2	2	2	2	2	2	2
3	1	1	1	1	2	1
4	1	2	1	1	2	1
5	1	2	1	1	1	1
6	1	2	2	1	2	1
7	2	2	1	1	2	1
8	1	2	1	1	1	1
9	1	1	2	2	1	1
10	1	2	1	1	2	2
11	2	2	2	2	2	2
12	2	2	1	1	2	2
13	2	2	1	2	1	2
14	2	2	1	2	2	2
15	1	1	2	1	1	1
16	1	1	1	1	2	2
17	2	2	1	1	2	2
18	1	1	1	1	1	1
19	1	2	1	2	1	2
20	2	2	1	1	2	1
21	1	2	1	2	1	1
22	1	2	1	1	1	1
23	1	2	1	1	1	2
24	2	2	1	2	1	2
25	1	2	1	2	2	1
26	1	2	1	1	2	2
27	2	2	2	1	1	2
28	1	2	1	1	1	1
29	2	2	1	2	1	2
30	2	1	1	1	1	2
31	2	1	1	1	1	1
32	2	2	2	1	2	2
33	1	2	1	1	2	2
34	1	2	1	1	1	1

Intraclass correlation - .6033

Pairwise Comparison - 0.56

Table 4.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 4 B2

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	M.4.10			2	M.4.20			1	M.4.10			1	M.4.10			2	M.4.2			1	M.4.10		
2	2	M.4.3			2	M.4.3			2	M.4.3			2	M.4.3			2	M.4.2			2	M.4.3	MHM1	
3	1	M.4.2			1	M.4.1			1	M.4.1			1	M.4.2			2	M.4.2			1	M.4.2	M.4.10	
4	1	M.4.12			2	M.4.12			1	M.4.12			1	M.4.12			2	M.4.13			1	M.4.12		
5	1	M.4.25			2	M.4.25			1	M.4.25			1	M.4.25			1	M.4.25			1	M.4.25	M.4.9	
6	1	M.4.2			2	M.4.20			2	M.4.20			1	M.4.20			2	M.4.2			1	M.4.20		
7	2	M.4.18			2	M.4.18			1	M.4.18			1	M.4.18			2	M.4.18			1	M.4.18		
8	1	M.4.22			2	M.4.22			1	M.4.22			1	M.4.22			1	M.4.22			1	M.4.22		
9	1	NBT.NF.M.4A			1	NBT.NF.M.4A			2	NBT.NF.M.4A			2	NBT.NF.M.4A			1	M.4.6			1	NBT.NF.M.4A		
10	1	M.4.4			2	M.4.4			1	M.4.7			1	M.4.4			2	M.4.2			2	M.4.4		
11	2	M.4.11			2	M.4.11			2	M.4.11			2	M.4.11			2	M.4.11			2	M.4.11	MHM4	
12	2	M.4.21			2	M.4.21			1	M.4.10			1	M.4.21			2	M.4.21			2	M.4.21		
13	2	M.4.13			2	M.4.13			1	M.4.13			2	M.4.13			1	M.4.12			2	M.4.13		
14	2	M.4.3			2	M.4.20			1	M.4.3			2	M.4.3			2	M.4.2			2	M.4.3	MHM1	
15	1	M.4.28			1	M.4.28			2	M.4.26			1	M.4.28			1	M.4.28			1	M.4.28		
16	1	M.4.5			1	M.4.5			1	M.4.5			1	M.4.5			2	M.4.5			2	M.4.5		
17	2	M.4.18			2	M.4.18			1	M.4.18			1	M.4.18			2	M.4.18			2	M.4.18		
18	1	M.4.8			1	M.4.8			1	M.4.8			1	M.4.8			1	M.4.8			1	M.4.8		
19	1	M.4.2			2	M.4.20			1	M.4.2			2	M.4.2			1	M.4.2			2	M.4.2		
20	2	M.4.17			2	M.4.14			1	M.4.16			1	M.4.17			2	M.4.16			1	M.4.17		
21	1	M.4.10			2	M.4.2			1	M.4.10			2	M.4.10			1	M.4.9			1	M.4.10		
22	1	M.4.24			2	M.4.24			1	M.4.24			1	M.4.24			1	M.4.24			1	M.4.24		
23	1	M.4.15			2	M.4.15			1	M.4.14			1	M.4.15			1	M.4.14			2	M.4.15	MHM2	
24	2	M.4.26			2	M.4.26			1	M.4.26			2	M.4.26			1	M.4.26			2	M.4.26		
25	1	M.4.10			2	M.4.1			1	M.4.10			2	M.4.7			2	M.4.10			1	M.4.10		
26	1	M.4.4			2	M.4.4			1	M.4.1			1	M.4.4			2	M.4.2			2	M.4.4		
27	2	M.4.16			2	M.4.16			2	M.4.16			1	M.4.16			1	M.4.16			2	M.4.16		
28	1	M.4.27			2	M.4.27			1	M.4.27			1	M.4.27			1	M.4.27			1	MD.G.M.4D		
29	2	M.4.1			2	M.4.20			1	M.4.1			2	M.4.1			1	M.4.2			2	M.4.1		
30	2	M.4.9			1	M.4.9			1	M.4.9			1	M.4.9			1	M.4.9			2	M.4.9		
31	2	M.4.17			1	M.4.16			1	M.4.17			1	M.4.17	M.4.16		1	M.4.18			1	M.4.17		
32	2	M.4.14			2	M.4.14			2	M.4.14			1	M.4.14			2	M.4.14			2	M.4.14		
33	1	M.4.19			2	M.4.21			1	M.4.19			1	M.4.19			2	M.4.20			2	M.4.19		
34	1	M.4.5			2	M.4.5			1	M.4.5			1	M.4.5			1	M.4.5			1	M.4.5		
Objective Pairwise Comparison: 0.66																								
Standard Pairwise Comparison: 0.86																								

Table 4.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers) WV MATH 2019 Grade 4 B2

Low		Medium				High				
0		7.2				12				
OA.M.4										
OA.M.4A										
M.4.1	3(2)	26(1)	25(1)	29(4)						
M.4.2	29(1)	26(1)	19(5)	21(1)	3(4)	1(1)	2(1)	6(2)	10(1)	14(1)
M.4.3	14(4)	2(5)								
OA.M.4B										
M.4.4	10(4)	26(4)								
OA.M.4C										
M.4.5	16(6)	34(6)								
NBT.NF.M.4										
NBT.NF.M.4A	9(5)									
M.4.6	9(1)									
M.4.7	10(1)	25(1)								
M.4.8	18(6)									
NBT.NF.M.4B										
M.4.9	21(1)	30(6)	5(1)							
M.4.10	25(4)	21(4)	12(1)	1(4)	3(1)					
M.4.11	11(12)									
NBT.NF.M.4C										
M.4.12	13(1)	4(5)								
M.4.13	4(1)	13(5)								
NBT.NF.M.4D										
M.4.14	23(2)	20(1)	32(6)							
M.4.15	23(4)									

NBT.NF.M.4E										
M.4.16	20(2)	27(6)	31(2)							
M.4.17	31(4)	20(3)								
M.4.18	31(1)	17(6)	7(6)							
MD.G.M.4										
MD.G.M.4A										
M.4.19	33(4)									
M. 4.20	33(1)	29(1)	1(1)	14(1)	6(4)	19(1)				
M.4.21	12(5)	33(1)								
MD.G.M.4B										
M.4.22	8(6)									
MD.G.M.4C										
M.4.23										
M.4.24	22(6)									
M.4.25	5(6)									
MD.G.M.4D	28(1)									
M.4.26	24(6)	15(1)								
M.4.27	28(5)									
M.4.28	15(5)									
MHM										
MHM1	14(1)	2(1)								
MHM2	23(1)									
MHM3										
MHM4	11(2)									
MHM5										
MHM6										
MHM7										
MHM8										

Table 4.8 *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)*
WV MATH 2019 Grade 4 B2

	Low	Medium	High	
	2.4	7.2		12
1 10199-8363	M.4.2:1	M.4.10:4	M. 4.20:1	
2 10199-894	M.4.2:1	M.4.3:5	MHM1:1	
3 10199-383	M.4.1:2	M.4.2:4	M.4.10:1	
4 REP10199-9890	M.4.12:5	M.4.13:1		
5 10199-892	M.4.9:1	M.4.25:6		
6 10199-768	M.4.2:2	M. 4.20:4		
7 REP10199-879	M.4.18:6			
8 REP10199-962	M.4.22:6			
9 10199-557	NBT.NF.M.4A:5	M.4.6:1		
10 10199-11504	M.4.2:1	M.4.4:4	M.4.7:1	
11 REP10199-9728	M.4.11:12	MHM4:2		
12 10199-11410	M.4.10:1	M.4.21:5		
13 10199-11419	M.4.12:1	M.4.13:5		
14 REP10199-678	M.4.2:1	M.4.3:4	M. 4.20:1	MHM1:1
15 10199-11474	M.4.26:1	M.4.28:5		
16 REP10199-729	M.4.5:6			
17 REP10199-881	M.4.18:6			
18 10199-9757	M.4.8:6			
19 REP10199-553	M.4.2:5	M. 4.20:1		
20 10199-791	M.4.14:1	M.4.16:2	M.4.17:3	
21 10199-124	M.4.2:1	M.4.9:1	M.4.10:4	
22 REP10199-963	M.4.24:6			
23 REP10199-9822	M.4.14:2	M.4.15:4	MHM2:1	
24 10199-303	M.4.26:6			
25 REP10199-81	M.4.1:1	M.4.7:1	M.4.10:4	
26 REP10199-11526	M.4.1:1	M.4.2:1	M.4.4:4	
27 10199-11417	M.4.16:6			
28 10199-11245	MD.G.M.4D:1	M.4.27:5		
29 10199-9793	M.4.1:4	M.4.2:1	M. 4.20:1	
30 10199-10055	M.4.9:6			
31 10199-687	M.4.16:2	M.4.17:4	M.4.18:1	
32 10199-721	M.4.14:6			
33 10199-10082	M.4.19:4	M. 4.20:1	M.4.21:1	
34 10199-835	M.4.5:6			

Table 4.9

Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])

WV MATH 2019 Grade 4 B2

Low DOK		Matched DOK		High DOK

OA.M.4										
OA.M.4A										
M.4.1: [2]	3:(2)[1]	25:(1)[2]	26:(1)[1]	29:(4)[2]						
M.4.2: [2]	1:(1)[2]	2:(1)[2]	3:(4)[1]	6:(2)[2]	10:(1)[2]	14:(1)[2]	19:(5)[1]	21:(1)[2]	26:(1)[2]	29:(1)[1]
M.4.3: [2]	2:(5)[2]	14:(4)[2]								
OA.M.4B										
M.4.4: [1]	10:(4)[2]	26:(4)[2]								
OA.M.4C										
M.4.5: [2]	16:(6)[1]	34:(6)[1]								
NBT.NF.M.4										
NBT.NF.M.4A: [2]	9:(5)[1]									
M.4.6: [1]	9:(1)[1]									
M.4.7: [1]	10:(1)[1]	25:(1)[2]								
M.4.8: [1]	18:(6)[1]									
NBT.NF.M.4B										
M.4.9: [1]	5:(1)[1]	21:(1)[1]	30:(6)[1]							
M.4.10: [2]	1:(4)[1]	3:(1)[1]	12:(1)[1]	21:(4)[1]	25:(4)[1]					
M.4.11: [2]	11:(12)[2]									
NBT.NF.M.4C										
M.4.12: [2]	4:(5)[1]	13:(1)[1]								
M.4.13: [2]	4:(1)[2]	13:(5)[2]								
NBT.NF.M.4D										
M.4.14: [2]	20:(1)[2]	23:(2)[1]	32:(6)[2]							
M.4.15: [2]	23:(4)[2]									
NBT.NF.M.4E										
M.4.16: [1]	20:(2)[2]	27:(6)[2]	31:(2)[1]							
M.4.17: [1]	20:(3)[1]	31:(4)[1]								
M.4.18: [2]	7:(6)[2]	17:(6)[2]	31:(1)[1]							
MD.G.M.4										

MD.G.M.4A										
M.4.19: [1]	33:(4)[1]									
M.4.20: [2]	1:(1)[2]	6:(4)[2]	14:(1)[2]	19:(1)[2]	29:(1)[2]	33:(1)[2]				
M.4.21: [2]	12:(5)[2]	33:(1)[2]								
MD.G.M.4B										
M.4.22: [2]	8:(6)[1]									
MD.G.M.4C										
M.4.23										
M.4.24: [1]	22:(6)[1]									
M.4.25: [2]	5:(6)[1]									
MD.G.M.4D: [2]	28:(1)[1]									
M.4.26: [1]	15:(1)[2]	24:(6)[2]								
M.4.27: [1]	28:(5)[1]									
M.4.28: [1]	15:(5)[1]									
MHM										
MHM1: [3]	2:(1)[2]	14:(1)[2]								
MHM2: [3]	23:(1)[2]									
MHM3										
MHM4: [3]	11:(2)[2]									
MHM5										
MHM6										
MHM7										
MHM8										

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 4B2 MATH

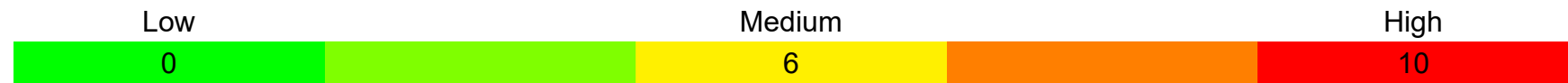
[illegible]

Table 4. CONF Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers) CONF WV Grade 4B2 MATH

Low		Medium		High	
2		6		10	
1	10199-8363	Exact:6			
2	10199-894	Exact:6			
3	10199-383	Exact:6			
4	REP10199-9890	Exact:6			
5	10199-892	Exact:4	Partial:2		
6	10199-768	Exact:3	Partial:3		
7	REP10199-879	Exact:5	Partial:1		
8	REP10199-962	Exact:5	Partial:1		
9	10199-557	Exact:3	Partial:1	Negligible:2	
10	10199-11504	Exact:4	Partial:2		
11	REP10199-9728	Exact:10	Partial:2		
12	10199-11410	Partial:4	Minimal:2		
13	10199-11419	Exact:6			
14	REP10199-678	Exact:6			
15	10199-11474	Exact:6			
16	REP10199-729	Exact:6			
17	REP10199-881	Exact:6			
18	10199-9757	Exact:6			
19	REP10199-553	Exact:6			
20	10199-791	Exact:4	Partial:2		
21	10199-124	Exact:5	Partial:1		
22	REP10199-963	Exact:6			
23	REP10199-9822	Exact:5	Partial:1		
24	10199-303	Exact:5	Partial:1		
25	REP10199-81	Exact:5	Partial:1		
26	REP10199-11526	Exact:5	Partial:1		
27	10199-11417	Exact:5	Partial:1		
28	10199-11245	Exact:3	Partial:2	Negligible:1	
29	10199-9793	Exact:6			
30	10199-10055	Exact:5	Partial:1		
31	10199-687	Exact:5	Partial:1		
32	10199-721	Exact:5	Partial:1		
33	10199-10082	Exact:6			
34	10199-835	Exact:6			

Batch 4 West Virginia Math Grade 4

Table 4.1 *Categorical Concurrence between Standards and Assessment as Rated by Three Reviewers WV MATH 2019 Grade 4 B4 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.4 Operations and Algebraic Thinking (OA)	3	5	1 2	1 4	20 80	7.33	1.15	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	5	13	1 2	6 7	46.15 53.85	17.33	1.15	YES
MD.G.M.4 Measurement and Data and Geometry	4	10	1 2	6 4	60 40	10.33	2.31	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	20	36	1 2 3	13 17 6	36 47 17	34.99	0	

Table 4.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Three Reviewers WV MATH 2019 Grade 4 B4 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.4 Operations and Algebraic Thinking (OA)	3	5	7.33	1.15	8.33	7	77.78	5	13.89	2	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	5	13	17.33	1.15	37.73	18	35.65	26	26.62	7	YES
MD.G.M.4 Measurement and Data and Geometry	4	10	10.33	2.31	7.41	6	72.36	9	20.23	16	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	20	36	34.99	0	22.86	13.1	54.29	11.4	22.86	2.9	

Table 4.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Three Reviewers WV MATH 2019 Grade 4 B4 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.4 Operations and Algebraic Thinking (OA)	3	5	7.33	1.15	4.33	0.58	86.67	11.55	YES	22	3	0.85	0.02	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	5	13	17.33	1.15	11	1.73	84.62	13.32	YES	48	3	0.83	0.01	YES
MD.G.M.4 Measurement and Data and Geometry	4	10	10.33	2.31	7.67	0.58	76.67	5.77	YES	30	7	0.82	0.1	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	20	36	34.99	0	5.8	4.7	61.99	42		25	20	0.83	0.08	

Table 4.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Three Reviewers WV MATH 2019 Grade 4 B4 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.4 Operations and Algebraic Thinking (OA)	YES	YES	YES	YES
NBT.NF.M.4 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.4 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 4.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 4 B4

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	1	2		1		
2	2	2		2		
3	2	2		1		
4	1	2		1		
5	2	2		2		
6	2	3		2		
7	2	2		2		
8	1	1		1		
9	1	2		1		
10	1	1		1		
11	2	2		2		
12	2	1		2		
13	1	1		1		
14	2	2		1		
15	2	2		1		
16	2	2		2		
17	1	2		1		
18	1	2		1		
19	2	2		2		
20	2	2		2		
21	1	2		1		
22	2	2		2		
23	2	2		2		
24	1	1		1		
25	2	2		2		
26	2	2		2		
27	1	2		1		
28	1	1		1		
29	2	2		2		
30	2	2		2		
31	2	2		2		
32	1	2		1		
33	2	2		2		
34	1	1		1		

Intraclass correlation - .8667

Pairwise Comparison - 0.75

Table 4.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 4 B4 Number of Reviewers: Three

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	M.4.1			2	M.4.2			1	M.4.1		
2	2	M.4.1			2	M. 4.20			2	M.4.1		
3	2	M.4.18			2	M.4.18			1	M.4.18		
4	1	M.4.13			2	M.4.12			1	M.4.12		
5	2	M.4.21			2	M.4.21			2	M.4.21		
6	2	M.4.25			3	M.4.25			2	M.4.25		
7	2	M.4.3			2	M. 4.20			2	M.4.3		
8	1	M.4.10			1	M.4.10			1	M.4.10		
9	1	M.4.22			2	M.4.22			1	M.4.22		
10	1	M.4.11			1	M.4.11			1	M.4.11		
11	2	M. 4.20			2	M. 4.20			2	M. 4.20		
12	2	M.4.8			1	M.4.8			2	M.4.8		
13	1	M.4.4			1	M.4.4			1	M.4.4		
14	2	M.4.18			2	M.4.18			1	M.4.18		
15	2	M.4.13			2	M.4.13			1	M.4.13		
16	2	M.4.1			2	M. 4.20			2	M.4.2		
17	1	M.4.27			2	M.4.27			1	M.4.27		
18	1	M.4.11			2	M.4.1			1	M.4.6		
19	2	M.4.5			2	M.4.5			2	M.4.5		
20	2	M.4.17			2	M.4.17			2	M.4.17		
21	1	M.4.10			2	M.4.10			1	M.4.10		
22	2	M. 4.20			2	M.4.19			2	M. 4.20		
23	2	M.4.3			2	M. 4.20			2	M.4.3		
24	1	M.4.28			1	M.4.28			1	M.4.28		
25	2	M.4.10			2	M.4.1			2	M.4.7		
26	2	M.4.17			2	M.4.17			2	M.4.17		
27	1	M.4.24			2	M.4.24			1	M.4.24		
28	1	M.4.9			1	M.4.9			1	M.4.9		
29	2	M.4.4			2	M.4.4			2	M.4.4		
30	2	M.4.16			2	M.4.14			2	M.4.16		
31	2	M.4.8			2	M.4.8			2	M.4.8		
32	1	M.4.15			2	M.4.15			1	M.4.15		
33	2	M.4.27			2	M.4.27			2	M.4.26		
34	1	M.4.14			1	M.4.14			1	M.4.14		
Objective Pairwise Comparison: 0.75												
Standard Pairwise Comparison: 0.88												

Table 4.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers) WV MATH 2019 Grade 4 B4

Low		Medium		High	
0		3.6		6	
OA.M.4					
OA.M.4A					
M.4.1	1(2)	2(2)	16(1)	25(1)	18(1)
M.4.2	16(1)	1(1)			
M.4.3	7(2)	23(2)			
OA.M.4B					
M.4.4	29(3)	13(3)			
OA.M.4C					
M.4.5	19(3)				
NBT.NF.M.4					
NBT.NF.M.4A					
M.4.6	18(1)				
M.4.7	25(1)				
M.4.8	12(3)	31(3)			
NBT.NF.M.4B					
M.4.9	28(3)				
M.4.10	25(1)	21(3)	8(3)		
M.4.11	10(3)	18(1)			
NBT.NF.M.4C					
M.4.12	4(2)				
M.4.13	4(1)	15(6)			
NBT.NF.M.4D					
M.4.14	30(1)	34(3)			

M.4.15	32(3)					
NBT.NF.M.4E						
M.4.16	30(2)					
M.4.17	26(3)	20(3)				
M.4.18	14(3)	3(3)				
MD.G.M.4						
MD.G.M.4A						
M.4.19	22(1)					
M. 4.20	22(2)	23(1)	16(1)	7(1)	2(1)	11(3)
M.4.21	5(3)					
MD.G.M.4B						
M.4.22	9(3)					
MD.G.M.4C						
M.4.23						
M.4.24	27(3)					
M.4.25	6(3)					
MD.G.M.4D						
M.4.26	33(1)					
M.4.27	33(2)	17(3)				
M.4.28	24(3)					
MHM						
MHM1						
MHM2						
MHM3						
MHM4						
MHM5						
MHM6						
MHM7						
MHM8						

Table 4.8 *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers) WV MATH 2019 Grade 4 B4*

Low		Medium		High	
1.2		3.6		6	
1	10199-447	M.4.1:2	M.4.2:1		
2	10199-824	M.4.1:2	M. 4.20:1		
3	10199-863	M.4.18:3			
4	REP10199-9890	M.4.12:2	M.4.13:1		
5	10199-680	M.4.21:3			
6	10199-874	M.4.25:3			
7	10199-887	M.4.3:2	M. 4.20:1		
8	REP10199-1289	M.4.10:3			
9	10199-11494	M.4.22:3			
10	10199-82	M.4.11:3			
11	10199-782	M. 4.20:3			
12	10199-606	M.4.8:3			
13	10199-716	M.4.4:3			
14	REP10199-881	M.4.18:3			
15	10199-1259	M.4.13:6			
16	10199-588	M.4.1:1	M.4.2:1	M. 4.20:1	
17	10199-235	M.4.27:3			
18	10199-860	M.4.1:1	M.4.6:1	M.4.11:1	
19	10199-731	M.4.5:3			
20	10199-789	M.4.17:3			
21	10199-391	M.4.10:3			
22	10199-1047	M.4.19:1	M. 4.20:2		
23	10199-7043	M.4.3:2	M. 4.20:1		
24	10199-86	M.4.28:3			
25	REP10199-81	M.4.1:1	M.4.7:1	M.4.10:1	
26	10199-793	M.4.17:3			
27	REP10199-963	M.4.24:3			
28	10199-697	M.4.9:3			
29	10199-10054	M.4.4:3			
30	10199-811	M.4.14:1	M.4.16:2		
31	10199-719	M.4.8:3			
32	REP10199-9854	M.4.15:3			
33	10199-613	M.4.26:1	M.4.27:2		
34	REP10199-9920	M.4.14:3			

Table 4.9

Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])

WV MATH 2019 Grade 4 B4

Low DOK		Matched DOK		High DOK

OA.M.4						
OA.M.4A						
M.4.1: [2]	1:(2)[1]	2:(2)[2]	16:(1)[2]	18:(1)[2]	25:(1)[2]	
M.4.2: [2]	1:(1)[2]	16:(1)[2]				
M.4.3: [2]	7:(2)[2]	23:(2)[2]				
OA.M.4B						
M.4.4: [1]	13:(3)[1]	29:(3)[2]				
OA.M.4C						
M.4.5: [2]	19:(3)[2]					
NBT.NF.M.4						
NBT.NF.M.4A						
M.4.6: [1]	18:(1)[1]					
M.4.7: [1]	25:(1)[2]					
M.4.8: [1]	12:(3)[2]	31:(3)[2]				
NBT.NF.M.4B						
M.4.9: [1]	28:(3)[1]					
M.4.10: [2]	8:(3)[1]	21:(3)[1]	25:(1)[2]			
M.4.11: [2]	10:(3)[1]	18:(1)[1]				
NBT.NF.M.4C						
M.4.12: [2]	4:(2)[2]					
M.4.13: [2]	4:(1)[1]	15:(6)[2]				
NBT.NF.M.4D						
M.4.14: [2]	30:(1)[2]	34:(3)[1]				
M.4.15: [2]	32:(3)[1]					
NBT.NF.M.4E						
M.4.16: [1]	30:(2)[2]					
M.4.17: [1]	20:(3)[2]	26:(3)[2]				
M.4.18: [2]	3:(3)[2]	14:(3)[2]				

MD.G.M.4						
MD.G.M.4A						
M.4.19: [1]	22:(1)[2]					
M. 4.20: [2]	2:(1)[2]	7:(1)[2]	11:(3)[2]	16:(1)[2]	22:(2)[2]	23:(1)[2]
M.4.21: [2]	5:(3)[2]					
MD.G.M.4B						
M.4.22: [2]	9:(3)[1]					
MD.G.M.4C						
M.4.23						
M.4.24: [1]	27:(3)[1]					
M.4.25: [2]	6:(3)[2]					
MD.G.M.4D						
M.4.26: [1]	33:(1)[2]					
M.4.27: [1]	17:(3)[1]	33:(2)[2]				
M.4.28: [1]	24:(3)[1]					
MHM						
MHM1						
MHM2						
MHM3						
MHM4						
MHM5						
MHM6						
MHM7						
MHM8						

Batch 1 West Virginia Math Grade 5

Table 5.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 5 B1 - Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.5 Operations and Algebraic Thinking (OA)	2	3	1 2	1 2	33.33 66.67	10.17	0.41	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	4	14	1 2	4 10	28.57 71.43	15.83	0.41	YES
MD.G.M.5 Measurement and Data and Geometry	5	9	1 2	4 5	44.44 55.56	10	0	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	1.5	1.22	NO
Total	19	34	1 2 3	9 19 6	26 56 18	37.5	1.22	

Table 5.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 5 B1 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.5 Operations and Algebraic Thinking (OA)	2	3	10.17	0.41	21.06	12	47.58	19	31.36	14	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	4	14	15.83	0.41	48.33	11	42.08	12	9.58	7	YES
MD.G.M.5 Measurement and Data and Geometry	5	9	10	0	43.33	20	50	21	6.67	10	YES
MHM Mathematical Habits of Mind	8	8	1.5	1.22	100	0	0	0	0	0	NO
Total	19	34	37.5	1.22	41.78	11.5	44	13.4	14.22	5.4	

Table 5.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 5 B1 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.5 Operations and Algebraic Thinking (OA)	2	3	10.17	0.41	3	0	100	0	YES	23	1	0.86	0.04	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	4	14	15.83	0.41	10.67	0.82	76.19	5.83	YES	45	1	0.83	0	YES
MD.G.M.5 Measurement and Data and Geometry	5	9	10	0	6.5	0.55	72.22	6.09	YES	28	0	0.84	0.03	YES
MHM Mathematical Habits of Mind	8	8	1.5	1.22	1	0	12.5	0	NO	3	1	1	0	YES
Total	19	34	37.5	1.22	5.3	4.24	65.23	37		25	17	0.88	0.08	

Table 5.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 5 B1 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.5 Operations and Algebraic Thinking (OA)	YES	YES	YES	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.5 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NO	NO	YES

Table 5.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 5 B1

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	2	2	2	2	2
2	2	2	2	2	2	2
3	1	1	1	1	1	1
4	1	1	1	1	1	1
5	1	1	1	1	1	1
6	2	2	1	2	2	1
7	1	1	1	1	2	1
8	2	2	2	2	1	2
9	1	1	1	1	1	1
10	1	1	1	1	1	1
11	1	2	1	1	2	1
12	1	2	2	2	1	2
13	1	2	1	2	1	1
14	2	1	1	1	1	2
15	1	1	1	2	2	2
16	1	2	1	1	2	1
17	1	1	1	1	1	1
18	1	1	1	1	1	1
19	1	2	1	1	1	1
20	2	2	2	2	1	2
21	1	2	1	1	1	1
22	2	2	2	2	2	2
23	1	2	1	1	1	1
24	1	2	1	1	1	1
25	1	1	1	1	1	1
26	1	2	1	2	1	1
27	2	1	1	2	2	2
28	1	1	1	1	1	1
29	2	1	1	2	2	2
30	2	2	1	1	2	2
31	1	1	1	1	1	1
32	1	1	2	1	2	1
33	2	2	1	1	2	1
34	1	2	1	1	2	1

Intraclass correlation - .8271

Pairwise Comparison - 0.71

Table 5.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 5 B1 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.5.12	MHM3		2	M.5.12	MHM3		2	M.5.12	MHM3		2	M.5.12	MHM3		2	M.5.12	MHM3		2	M.5.12		
2	2	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3		
3	1	M.5.19			1	M.5.19			1	M.5.19			1	M.5.19			1	M.5.19			1	M.5.19		
4	1	M.5.9			1	M.5.9			1	M.5.9			1	M.5.9			1	M.5.9			1	M.5.9		
5	1	M.5.13			1	M.5.13			1	M.5.13			1	M.5.13			1	M.5.13			1	M.5.13		
6	2	M.5.25			2	M.5.25			1	M.5.25			2	M.5.25			2	M.5.25			1	M.5.26		
7	1	M.5.23			1	M.5.23			1	M.5.24			1	M.5.24			2	M.5.24			1	M.5.23		
8	2	M.5.1			2	M.5.1			2	M.5.1			2	M.5.1			1	M.5.1			2	M.5.1	MHM3	
9	1	M.5.10			1	M.5.10			1	M.5.10			1	M.5.10			1	M.5.2			1	M.5.10		
10	1	M.5.7			1	M.5.7			1	M.5.7			1	M.5.7			1	M.5.7			1	M.5.7		
11	1	M.5.16			2	M.5.16			1	M.5.16			1	M.5.16			2	M.5.16			1	M.5.16		
12	1	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3			1	M.5.3			2	M.5.3		
13	1	M.5.21			2	M.5.21			1	M.5.21			2	M.5.21			1	M.5.21			1	M.5.22		
14	2	M.5.2			1	M.5.1			1	M.5.1			1	M.5.1			1	M.5.1			2	M.5.1		
15	1	M.5.11			1	M.5.11			1	M.5.11			2	M.5.11			2	M.5.11			2	M.5.11		
16	1	M.5.18			2	M.5.18			1	M.5.18			1	M.5.18			2	M.5.18			1	M.5.18		
17	1	M.5.13			1	M.5.13			1	M.5.13			1	M.5.13			1	M.5.13			1	M.5.13		
18	1	M.5.10			1	M.5.10			1	M.5.10			1	M.5.10			1	M.5.10			1	M.5.10		
19	1	M.5.2			2	M.5.2			1	M.5.2			1	M.5.2			1	M.5.2			1	M.5.2		
20	2	M.5.24			2	M.5.24			2	M.5.24			2	M.5.24			1	M.5.23			2	M.5.24		
21	1	M.5.9			2	M.5.8			1	M.5.9			1	M.5.9			1	M.5.9			1	M.5.9		
22	2	M.5.3			2	M.5.3			2	M.5.2			2	M.5.3			2	M.5.3			2	M.5.3		
23	1	M.5.19			2	M.5.23			1	M.5.19			1	M.5.19			1	M.5.19			1	M.5.19		
24	1	M.5.2			2	M.5.2			1	M.5.2			1	M.5.2			1	M.5.2			1	M.5.2		
25	1	M.5.8			1	M.5.8			1	M.5.8			1	M.5.8			1	M.5.8			1	M.5.8		
26	1	M.5.21			2	M.5.22			1	M.5.21			2	M.5.21			1	M.5.21			1	M.5.22		
27	2	M.5.11			1	M.5.11			1	M.5.11			2	M.5.11			2	M.5.11			2	M.5.11		
28	1	M.5.7			1	M.5.7			1	M.5.7			1	M.5.7			1	M.5.7			1	M.5.7		
29	2	M.5.1			1	M.5.1			1	M.5.1			2	M.5.1			2	M.5.1			2	M.5.1	MHM3	
30	2	M.5.18			2	M.5.18			1	M.5.18			1	M.5.18			2	M.5.18			2	M.5.18		
31	1	M.5.6			1	M.5.6			1	M.5.6			1	M.5.6			1	M.5.10			1	M.5.6		
32	1	M.5.22			1	M.5.22			2	M.5.22			1	M.5.22			2	M.5.22			1	M.5.22		
33	2	M.5.14			2	M.5.14			1	M.5.14			1	M.5.14			2	M.5.11			1	M.5.14		
34	1	M.5.5			2	M.5.5			1	M.5.5			1	M.5.5			2	M.5.5			1	M.5.5		
Objective Pairwise Comparison: 0.84																								
Standard Pairwise Comparison: 0.96																								

Table 5.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV MATH 2019 Grade 5 B1

Low	Medium		High	
0	7.2		12	

OA.M.5					
OA.M.5A					
M.5.1	8(12)	14(5)	29(12)		
M.5.2	24(6)	19(6)	22(1)	14(1)	9(1)
OA.M.5B					
M.5.3	2(6)	22(5)	12(6)		
NBT.NF.M.5					
NBT.NF.M.5A					
M.5.4					
M.5.5	34(6)				
M.5.6	31(5)				
M.5.7	28(6)	10(6)			
NBT.NF.M.5B					
M.5.8	25(6)	21(1)			
M.5.9	21(5)	4(6)			
M.5.10	9(5)	18(6)	31(1)		
NBT.NF.M.5C					
M.5.11	33(1)	27(6)	15(6)		
M.5.12	1(6)				
NBT.NF.M.5D					
M.5.13	5(6)	17(6)			
M.5.14	33(5)				

M.5.15					
M.5.16	11(6)				
M.5.17					
MD.G.M.5					
MD.G.M.5A					
M.5.18	16(6)	30(6)			
MD.G.M.5B					
M.5.19	23(5)	3(6)			
MD.G.M.5C					
M.5.20					
M.5.21	13(5)	26(4)			
M.5.22	26(2)	13(1)	32(6)		
MD.G.M.5D					
M.5.23	23(1)	20(1)	7(3)		
M.5.24	7(3)	20(5)			
MD.G.M.5E					
M.5.25	6(5)				
M.5.26	6(1)				
MHM					
MHM1					
MHM2					
MHM3	8(2)	29(2)	1(5)		
MHM4					
MHM5					
MHM6					
MHM7					
MHM8					

Table 5.8 *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)*
WV MATH 2019 Grade 5 B1

Low	Medium	High
2.4	7.2	12

1 10199-1655	M.5.12:6	MHM3:5
2 10199-10047	M.5.3:6	
3 10199-1284	M.5.19:6	
4 10199-987	M.5.9:6	
5 10199-1714	M.5.13:6	
6 10199-1123	M.5.25:5	M.5.26:1
7 10199-1118	M.5.23:3	M.5.24:3
8 REP10199-1315	M.5.1:12	MHM3:2
9 10199-1087	M.5.2:1	M.5.10:5
10 10199-9765	M.5.7:6	
11 10199-11158	M.5.16:6	
12 10199-1068	M.5.3:6	
13 REP10199-7468	M.5.21:5	M.5.22:1
14 10199-10044	M.5.1:5	M.5.2:1
15 10199-1643	M.5.11:6	
16 REP10199-970	M.5.18:6	
17 10199-11347	M.5.13:6	
18 10199-1090	M.5.10:6	
19 10199-1179	M.5.2:6	
20 10199-1206	M.5.23:1	M.5.24:5
21 10199-1042	M.5.8:1	M.5.9:5
22 10199-1151	M.5.2:1	M.5.3:5
23 REP10199-1102	M.5.19:5	M.5.23:1
24 10199-1053	M.5.2:6	
25 10199-983	M.5.8:6	
26 REP10199-1763	M.5.21:4	M.5.22:2
27 10199-1657	M.5.11:6	
28 10199-1170	M.5.7:6	
29 10199-1314	M.5.1:12	MHM3:2
30 10199-2128	M.5.18:6	
31 10199-1005	M.5.6:5	M.5.10:1
32 10199-1458	M.5.22:6	
33 10199-1624	M.5.11:1	M.5.14:5
34 10199-1268	M.5.5:6	

Table 5.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
 WV MATH 2019 Grade 5 B1

Low DOK		Matched DOK		High DOK

OA.M.5					
OA.M.5A					
M.5.1: [1]	8:(12)[2]	14:(5)[1]	29:(12)[2]		
M.5.2: [2]	9:(1)[1]	14:(1)[2]	19:(6)[1]	22:(1)[2]	24:(6)[1]
OA.M.5B					
M.5.3: [2]	2:(6)[2]	12:(6)[2]	22:(5)[2]		
NBT.NF.M.5					
NBT.NF.M.5A					
M.5.4					
M.5.5: [2]	34:(6)[1]				
M.5.6: [1]	31:(5)[1]				
M.5.7: [1]	10:(6)[1]	28:(6)[1]			
NBT.NF.M.5B					
M.5.8: [1]	21:(1)[2]	25:(6)[1]			
M.5.9: [2]	4:(6)[1]	21:(5)[1]			
M.5.10: [2]	9:(5)[1]	18:(6)[1]	31:(1)[1]		
NBT.NF.M.5C					
M.5.11: [1]	15:(6)[2]	27:(6)[2]	33:(1)[2]		
M.5.12: [2]	1:(6)[2]				
NBT.NF.M.5D					
M.5.13: [2]	5:(6)[1]	17:(6)[1]			
M.5.14: [2]	33:(5)[1]				
M.5.15					
M.5.16: [2]	11:(6)[1]				
M.5.17					
MD.G.M.5					
MD.G.M.5A					
M.5.18: [2]	16:(6)[1]	30:(6)[2]			

MD.G.M.5B					
M.5.19: [2]	3:(6)[1]	23:(5)[1]			
MD.G.M.5C					
M.5.20					
M.5.21: [1]	13:(5)[1]	26:(4)[1]			
M.5.22: [2]	13:(1)[1]	26:(2)[2]	32:(6)[1]		
MD.G.M.5D					
M.5.23: [1]	7:(3)[1]	20:(1)[1]	23:(1)[2]		
M.5.24: [2]	7:(3)[1]	20:(5)[2]			
MD.G.M.5E					
M.5.25: [2]	6:(5)[2]				
M.5.26: [1]	6:(1)[1]				
MHM					
MHM1					
MHM2					
MHM3: [3]	1:(5)[2]	8:(2)[2]	29:(2)[2]		
MHM4					
MHM5					
MHM6					
MHM7					
MHM8					

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 5B1 MATH

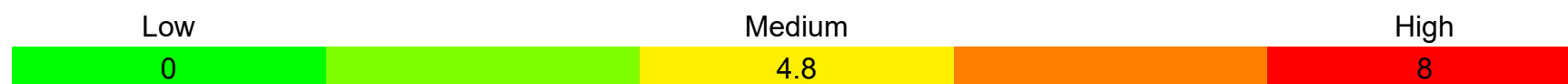
[illegible]

Table 5.CONF *Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)*
CONF WV Grade 5B1 MATH

Low		Medium		High	
1.6		4.8		8	
1	10199-1655	Exact:2	Partial:4		
2	10199-10047	Exact:5	Partial:1		
3	10199-1284	Exact:3	Partial:2	Minimal:1	
4	10199-987	Exact:6			
5	10199-1714	Exact:5	Partial:1		
6	10199-1123	Exact:5	Minimal:1		
7	10199-1118	Exact:4	Partial:1	Minimal:1	
8	REP10199-1315	Exact:6	Partial:6		
9	10199-1087	Exact:6			
10	10199-9765	Exact:6			
11	10199-11158	Exact:6			
12	10199-1068	Exact:6			
13	REP10199-7468	Exact:5	Minimal:1		
14	10199-10044	Exact:4	Partial:1	Minimal:1	
15	10199-1643	Exact:5	Partial:1		
16	REP10199-970	Exact:6			
17	10199-11347	Exact:5	Partial:1		
18	10199-1090	Exact:6			
19	10199-1179	Exact:6			
20	10199-1206	Exact:6			
21	10199-1042	Exact:6			
22	10199-1151	Exact:6			
23	REP10199-1102	Exact:4	Partial:1	Minimal:1	
24	10199-1053	Exact:6			
25	10199-983	Exact:5	Partial:1		
26	REP10199-1763	Exact:5	Minimal:1		
27	10199-1657	Exact:6			
28	10199-1170	Exact:6			
29	10199-1314	Exact:8	Partial:4		
30	10199-2128	Exact:6			
31	10199-1005	Exact:6			
32	10199-1458	Exact:6			
33	10199-1624	Exact:5	Partial:1		
34	10199-1268	Exact:4	Partial:2	Negligible:1	

Batch 2 West Virginia Math Grade 5

Table 5.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 5 B2 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.5 Operations and Algebraic Thinking (OA)	2	3	1 2	1 2	33.33 66.67	10.67	0.82	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	4	14	1 2	4 10	28.57 71.43	17.83	0.41	YES
MD.G.M.5 Measurement and Data and Geometry	5	9	1 2	4 5	44.44 55.56	10.5	0.84	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	1	1.67	NO
Total	19	34	1 2 3	9 19 6	26 56 18	40	1.67	

Table 5.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 5 B2 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.5 Operations and Algebraic Thinking (OA)	2	3	10.67	0.82	24.58	12	45.62	20	29.8	15	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	4	14	17.83	0.41	33.5	14	49.73	19	16.78	10	YES
MD.G.M.5 Measurement and Data and Geometry	5	9	10.5	0.84	32.9	19	53.36	18	13.74	12	YES
MHM Mathematical Habits of Mind	8	8	1	1.67	100	0	0	0	0	0	NO
Total	19	34	40	1.67	32.5	15.2	48.33	19.1	19.17	9.8	

Table 5.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 5 B2 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.5 Operations and Algebraic Thinking (OA)	2	3	10.67	0.82	3	0	100	0	YES	23	1	0.84	0.04	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	4	14	17.83	0.41	9.5	0.55	67.86	3.91	YES	46	1	0.85	0.02	YES
MD.G.M.5 Measurement and Data and Geometry	5	9	10.5	0.84	7	0.89	77.78	9.94	YES	30	2	0.81	0.03	YES
MHM Mathematical Habits of Mind	8	8	1	1.67	0.33	0.52	4.17	6.45	NO	1	2	1	0	YES
Total	19	34	40	1.67	5	4.08	62.45	41		25	18	0.88	0.08	

Table 5.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 5 B2 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.5 Operations and Algebraic Thinking (OA)	YES	YES	YES	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.5 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NO	NO	YES

Table 5.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 5 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	1	1	1	2	2
2	1	2	1	1	1	1
3	1	1	1	1	2	1
4	2	2	2	2	2	2
5	1	1	1	1	2	1
6	1	1	1	1	2	1
7	1	2	1	1	2	2
8	2	2	1	2	2	2
9	1	2	1	2	2	2
10	2	2	1	2	1	2
11	2	2	2	1	2	2
12	2	2	1	2	2	2
13	1	1	1	1	2	1
14	2	2	2	2	2	2
15	1	2	1	2	2	1
16	1	2	1	1	2	2
17	2	1	2	2	2	2
18	1	2	1	1	2	1
19	1	1	1	1	1	1
20	1	1	1	1	1	1
21	1	2	1	1	1	1
22	2	1	2	2	2	2
23	2	2	1	2	1	2
24	1	1	1	1	1	1
25	1	1	1	1	2	1
26	1	2	1	1	2	1
27	1	2	1	2	2	1
28	2	1	1	2	2	2
29	1	1	1	2	2	2
30	2	1	1	2	2	2
31	1	1	1	2	2	1
32	2	2	2	2	2	2
33	1	1	1	1	2	1
34	1	2	2	1	2	1

Intraclass correlation - .7739

Pairwise Comparison - 0.62

Table 5.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 5 B2 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.5.2			1	M.5.1			1	M.5.2			1	M.5.2			2	M.5.2			2	M.5.2		
2	1	M.5.19			2	M.5.23			1	M.5.19			1	M.5.19			1	M.5.19			1	M.5.19		
3	1	M.5.9			1	M.5.9			1	M.5.9			2	M.5.9			1	M.5.9						
4	2	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3		
5	1	M.5.12			1	M.5.11			1	M.5.11			2	M.5.12			1	M.5.12						
6	1	M.5.23			1	M.5.23			1	M.5.23			2	M.5.24			1	M.5.24			1	M.5.23		
7	1	M.5.16			2	M.5.16			1	M.5.16			2	M.5.16			2	M.5.16			2	M.5.16		
8	2	M.5.25			2	M.5.25			1	M.5.25			2	M.5.25			2	M.5.25			2	M.5.25		
9	1	M.5.16			2	M.5.10			1	M.5.10			2	M.5.10			2	M.5.10			2	M.5.10		
10	2	M.5.7			2	M.5.7			1	M.5.7			2	M.5.7			1	M.5.7			2	M.5.7		
11	2	M.5.13			2	M.5.13			2	M.5.13			1	M.5.13			2	M.5.13			2	M.5.13		
12	2	M.5.1			2	M.5.1			1	M.5.1			2	M.5.1			2	M.5.1			2	M.5.1	MHM3	
13	1	M.5.18			1	M.5.18			1	M.5.18			2	M.5.18			1	M.5.18			1	M.5.18		
14	2	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3			2	M.5.3		
15	1	M.5.12			2	M.5.18			1	M.5.11			2	M.5.12			2	M.5.12			1	M.5.12		
16	1	M.5.9			2	M.5.9			1	M.5.9			2	M.5.9			2	M.5.9			2	M.5.9		
17	2	M.5.22			1	M.5.21			2	M.5.22			2	M.5.22			2	M.5.22			2	M.5.22		
18	1	M.5.3			2	M.5.3			1	M.5.3			2	M.5.3			1	M.5.3			1	M.5.3		
19	1	M.5.13			1	M.5.13			1	M.5.13			1	M.5.13			1	M.5.13			1	M.5.13		
20	1	M.5.23			1	M.5.23			1	M.5.23			1	M.5.23			1	M.5.23			1	M.5.23		
21	1	M.5.10			2	M.5.10			1	M.5.10			1	M.5.10			1	M.5.10			1	M.5.10		
22	2	M.5.1			1	M.5.1			2	M.5.1			2	M.5.1			2	M.5.1			2	M.5.1		
23	2	M.5.11			2	M.5.10			1	M.5.11			2	M.5.11			1	M.5.13			2	M.5.11		
24	1	M.5.2			1	M.5.2			1	M.5.2			1	M.5.2			1	M.5.2			1	M.5.2		
25	1	M.5.7			1	M.5.7			1	M.5.7			2	M.5.7			1	M.5.7			1	M.5.7		
26	1	M.5.18			2	M.5.18			1	M.5.18			2	M.5.18			1	M.5.18			1	M.5.18		
27	1	M.5.21			2	M.5.22			1	M.5.21			2	M.5.21			2	M.5.21			1	M.5.22		
28	2	M.5.8			1	M.5.8			1	M.5.8			2	M.5.8			2	M.5.8			2	M.5.8		
29	1	M.5.1			1	M.5.1			1	M.5.1			2	M.5.1	MHM3		2	M.5.20			2	M.5.1	MHM3	
30	2	M.5.20			1	M.5.20			1	M.5.20			2	M.5.20			2	M.5.20			2	M.5.20		
31	1	M.5.11			1	M.5.11			1	M.5.11			2	M.5.11			2	M.5.11			1	M.5.11		
32	2	M.5.17			2	M.5.17			2	M.5.17			2	M.5.17			2	M.5.16			2	M.5.17		
33	1	M.5.5			1	M.5.5			1	M.5.5			2	M.5.5			1	M.5.5			1	M.5.5		
34	1	M.5.22			2	M.5.22			2	M.5.22			1	M.5.22			2	M.5.22			1	M.5.22		
Objective Pairwise Comparison: 0.84																								
Standard Pairwise Comparison: 0.97																								

Table 5.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV MATH 2019 Grade 5 B2

	Low	Medium	High
	0	7.2	12
OA.M.5			
OA.M.5A			
M.5.1	1(1)	12(12)	22(6)
M.5.2	24(6)	1(5)	
OA.M.5B			
M.5.3	4(6)	14(6)	18(12)
NBT.NF.M.5			
NBT.NF.M.5A			
M.5.4			
M.5.5	33(6)		
M.5.6			
M.5.7	25(6)	10(12)	
NBT.NF.M.5B			
M.5.8	28(6)		
M.5.9	16(6)	3(6)	
M.5.10	9(5)	23(1)	21(6)
NBT.NF.M.5C			
M.5.11	23(4)	15(1)	5(2)
M.5.12	5(4)	15(4)	
NBT.NF.M.5D			
M.5.13	11(6)	23(1)	19(6)
M.5.14			
M.5.15			

M.5.16	9(1)	7(6)	32(2)	
M.5.17	32(10)			
MD.G.M.5				
MD.G.M.5A				
M.5.18	15(1)	13(6)	26(6)	
MD.G.M.5B				
M.5.19	2(5)			
MD.G.M.5C				
M.5.20	29(2)	30(6)		
M.5.21	27(4)	17(1)		
M.5.22	17(5)	27(2)	34(6)	
MD.G.M.5D				
M.5.23	20(6)	2(1)	6(4)	
M.5.24	6(2)			
MD.G.M.5E				
M.5.25	8(6)			
M.5.26				
MHM				
MHM1				
MHM2				
MHM3	12(2)	29(4)		
MHM4				
MHM5				
MHM6				
MHM7				
MHM8				

Table 5.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 WV MATH 2019 Grade 5 B2

Low		Medium		High	
2.4		7.2		12	
1	10199-1040	M.5.1:1	M.5.2:5		
2	REP10199-1102	M.5.19:5	M.5.23:1		
3	10199-1051	M.5.9:6			
4	REP10199-7617	M.5.3:6			
5	10199-1593	M.5.11:2	M.5.12:4		
6	10199-1121	M.5.23:4	M.5.24:2		
7	10199-1704	M.5.16:6			
8	10199-1198	M.5.25:6			
9	10199-11350	M.5.10:5	M.5.16:1		
10	10199-1172	M.5.7:12			
11	REP10199-1736	M.5.13:6			
12	REP10199-1315	M.5.1:12	MHM3:2		
13	REP10199-968	M.5.18:6			
14	10199-1156	M.5.3:6			
15	10199-1623	M.5.11:1	M.5.12:4	M.5.18:1	
16	10199-10049	M.5.9:6			
17	REP10199-1000	M.5.21:1	M.5.22:5		
18	10199-1734	M.5.3:12			
19	REP10199-1712	M.5.13:6			
20	REP10199-1302	M.5.23:6			
21	10199-1530	M.5.10:6			
22	10199-1203	M.5.1:6			
23	REP10199-1663	M.5.10:1	M.5.11:4	M.5.13:1	
24	REP10199-11276	M.5.2:6			
25	10199-1288	M.5.7:6			
26	REP10199-970	M.5.18:6			
27	REP10199-1763	M.5.21:4	M.5.22:2		
28	10199-1021	M.5.8:6			
29	10199-1310	M.5.1:10	M.5.20:2	MHM3:4	
30	10199-1457	M.5.20:6			
31	10199-1573	M.5.11:6			
32	10199-1653	M.5.16:2	M.5.17:10		
33	10199-11285	M.5.5:6			
34	REP10199-1508	M.5.22:6			

Table 5.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV MATH 2019 Grade 5 B2

Low DOK		Matched DOK		High DOK

OA.M.5				
OA.M.5A				
M.5.1: [1]	1:(1)[1]	12:(12)[2]	22:(6)[2]	29:(10)[1]
M.5.2: [2]	1:(5)[2]	24:(6)[1]		
OA.M.5B				
M.5.3: [2]	4:(6)[2]	14:(6)[2]	18:(12)[1]	
NBT.NF.M.5				
NBT.NF.M.5A				
M.5.4				
M.5.5: [2]	33:(6)[1]			
M.5.6				
M.5.7: [1]	10:(12)[2]	25:(6)[1]		
NBT.NF.M.5B				
M.5.8: [1]	28:(6)[2]			
M.5.9: [2]	3:(6)[1]	16:(6)[2]		
M.5.10: [2]	9:(5)[2]	21:(6)[1]	23:(1)[2]	
NBT.NF.M.5C				
M.5.11: [1]	5:(2)[1]	15:(1)[1]	23:(4)[2]	31:(6)[1]
M.5.12: [2]	5:(4)[1]	15:(4)[2]		
NBT.NF.M.5D				
M.5.13: [2]	11:(6)[2]	19:(6)[1]	23:(1)[1]	
M.5.14				
M.5.15				
M.5.16: [2]	7:(6)[2]	9:(1)[1]	32:(2)[2]	
M.5.17: [2]	32:(10)[2]			
MD.G.M.5				
MD.G.M.5A				
M.5.18: [2]	13:(6)[1]	15:(1)[2]	26:(6)[1]	

MD.G.M.5B				
M.5.19: [2]	2:(5)[1]			
MD.G.M.5C				
M.5.20: [1]	29:(2)[2]	30:(6)[2]		
M.5.21: [1]	17:(1)[1]	27:(4)[2]		
M.5.22: [2]	17:(5)[2]	27:(2)[2]	34:(6)[2]	
MD.G.M.5D				
M.5.23: [1]	2:(1)[2]	6:(4)[1]	20:(6)[1]	
M.5.24: [2]	6:(2)[2]			
MD.G.M.5E				
M.5.25: [2]	8:(6)[2]			
M.5.26				
MHM				
MHM1				
MHM2				
MHM3: [3]	12:(2)[2]	29:(4)[2]		
MHM4				
MHM5				
MHM6				
MHM7				
MHM8				

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 5B2 MATH

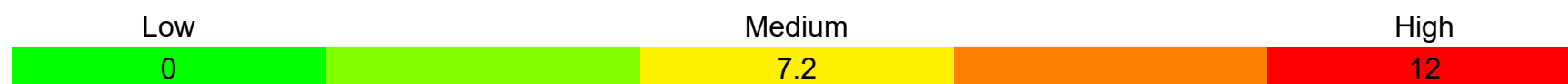
[illegible]

Table 5.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 5B2 MATH

Low		Medium		High	
2.4		7.2		12	
1	10199-1040	Exact:6			
2	REP10199-1102	Exact:4	Partial:1	Minimal:1	
3	10199-1051	Exact:6			
4	REP10199-7617	Exact:6			
5	10199-1593	Exact:6			
6	10199-1121	Exact:5	Negligible:1		
7	10199-1704	Exact:5	Partial:1		
8	10199-1198	Exact:6			
9	10199-11350	Exact:6			
10	10199-1172	Exact:12			
11	REP10199-1736	Exact:4	Partial:2		
12	REP10199-1315	Exact:10	Partial:2		
13	REP10199-968	Exact:6			
14	10199-1156	Exact:6			
15	10199-1623	Exact:6			
16	10199-10049	Exact:5	Partial:1		
17	REP10199-1000	Exact:6			
18	10199-1734	Exact:12			
19	REP10199-1712	Exact:6			
20	REP10199-1302	Exact:5	Partial:1		
21	10199-1530	Exact:6			
22	10199-1203	Exact:6			
23	REP10199-1663	Exact:6			
24	REP10199-11276	Exact:5	Partial:1		
25	10199-1288	Exact:6			
26	REP10199-970	Exact:6			
27	REP10199-1763	Exact:4	Minimal:2		
28	10199-1021	Exact:5	Partial:1		
29	10199-1310	Exact:10	Partial:2		
30	10199-1457	Exact:4	Partial:2		
31	10199-1573	Exact:6			
32	10199-1653	Exact:8	Partial:4		
33	10199-11285	Exact:5	Partial:1	Negligible:1	
34	REP10199-1508	Exact:4	Partial:1		

Batch 4 West Virginia Math Grade 5

Table 5.1 *Categorical Concurrence between Standards and Assessment as Rated by Three Reviewers WV MATH 2019 Grade 5 B4 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
OA.M.5 Operations and Algebraic Thinking (OA)	2	3	1 2	1 2	33.33 66.67	8	0	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	4	14	1 2	4 10	28.57 71.43	20.33	1.53	YES
MD.G.M.5 Measurement and Data and Geometry	5	9	1 2	4 5	44.44 55.56	9	1	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	19	34	1 2 3	9 19 6	26 56 18	37.33	1.15	

Table 5.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Three Reviewers WV MATH 2019 Grade 5 B4 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
OA.M.5 Operations and Algebraic Thinking (OA)	2	3	8	0	33.33	26	58.33	19	8.33	14	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	4	14	20.33	1.53	24.38	9	63.93	1	11.69	8	YES
MD.G.M.5 Measurement and Data and Geometry	5	9	9	1	22.5	20	66.3	20	11.2	1	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	19	34	37.33	1.15	25.89	11.9	63.39	8.5	10.71	6.9	

Table 5.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Three Reviewers WV MATH 2019 Grade 5 B4 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
OA.M.5 Operations and Algebraic Thinking (OA)	2	3	8	0	2.67	0.58	88.89	19.25	YES	23	0	0.79	0.11	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	4	14	20.33	1.53	9.67	0.58	69.05	4.12	YES	51	3	0.8	0.05	YES
MD.G.M.5 Measurement and Data and Geometry	5	9	9	1	7	0	77.78	0	YES	26	3	0.85	0.03	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	19	34	37.33	1.15	4.8	4.32	58.93	40		25	21	0.81	0.1	

Table 5.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Three Reviewers WV MATH 2019 Grade 5 B4 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
OA.M.5 Operations and Algebraic Thinking (OA)	YES	YES	YES	YES
NBT.NF.M.5 Number and Operations in Base Ten and Fractions	YES	YES	YES	YES
MD.G.M.5 Measurement and Data and Geometry	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 5.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 5 B4

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	1	2		2		
2	1	2		2		
3	1	1		1		
4	2	2		2		
5	2	2		2		
6	2	2		2		
7	2	1		1		
8	1	1		1		
9	1	1		2		
10	1	1		1		
11	2	3		2		
12	1	2		2		
13	2	2		2		
14	1	2		2		
15	2	1		2		
16	1	2		2		
17	2	2		2		
18	1	1		2		
19	1	2		2		
20	1	2		1		
21	1	1		1		
22	1	2		2		
23	2	2		2		
24	1	2		2		
25	1	1		1		
26	1	1		1		
27	2	2		1		
28	1	1		2		
29	1	1		1		
30	2	1		2		
31	2	2		2		
32	1	2		2		
33	2	2		2		
34	2	2		2		

Intraclass correlation - .7653

Pairwise Comparison - 0.65

Table 5.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 5 B4 Number of Reviewers: Three

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	M.5.14			2	M.5.14			2	M.5.15		
2	1	M.5.24			2	M.5.24			2	M.5.24		
3	1	M.5.19			1	M.5.23			1	M.5.19		
4	2	M.5.3			2	M.5.3			2	M.5.3		
5	2	M.5.10			2	M.5.10			2	M.5.10		
6	2	M.5.14	M.5.12		2	M.5.12			2	M.5.12	M.5.14	
7	2	M.5.26			1	M.5.26			1	M.5.25		
8	1	M.5.13			1	M.5.13			1	M.5.13		
9	1	M.5.2			1	M.5.2			2	M.5.2		
10	1	M.5.9			1	M.5.9			1	M.5.9		
11	2	M.5.14			3	M.5.14			2	M.5.14		
12	1	M.5.2			2	M.5.18			2	M.5.18		
13	2	M.5.3			2	M.5.2			2	M.5.1		
14	1	M.5.7			2	M.5.3			2	M.5.3		
15	2	M.5.22			1	M.5.7			2	M.5.7		
16	1	M.5.12			2	M.5.22			2	M.5.22		
17	2	M.5.10			2	M.5.12			2	M.5.12		
18	1	M.5.1			1	M.5.10			2	M.5.10		
19	1	M.5.25			2	M.5.2			2	M.5.1		
20	1	M.5.13			2	M.5.25			1	M.5.25		
21	1	M.5.2			1	M.5.13			1	M.5.13		
22	1	M.5.21			2	M.5.2			2	M.5.2		
23	2	M.5.12			2	M.5.21			2	M.5.21		
24	1	M.5.7			2	M.5.10			2	M.5.11		
25	1	M.5.23			1	M.5.7			1	M.5.7		
26	1	M.5.2			1	M.5.23			1	M.5.23		
27	2	M.5.8			2	M.5.2			1	M.5.1		
28	1	M.5.2			1	M.5.8			2	M.5.8		
29	1	M.5.10			1	M.5.2			1	M.5.2		
30	2	M.5.12			1	M.5.10			2	M.5.18		
31	2	M.5.12			2	M.5.11			2	M.5.11		
32	1	M.5.4			2	M.5.4			2	M.5.10		
33	2	M.5.16			2	M.5.16			2	M.5.17		
34	2	M.5.22			2	M.5.22			2	M.5.22		
Objective Pairwise Comparison: 0.45												
Standard Pairwise Comparison: 0.69												

Table 5.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers) WV MATH 2019 Grade 5 B4

	Low		Medium				High			
	0		3.6				6			
OA.M.5										
OA.M.5A										
M.5.1	13(1)	18(1)	19(1)	27(1)						
M.5.2	27(1)	28(1)	29(2)	26(1)	19(1)	21(1)	22(2)	12(1)	13(1)	9(3)
OA.M.5B										
M.5.3	13(1)	14(2)	4(3)							
NBT.NF.M.5										
NBT.NF.M.5A										
M.5.4	32(2)									
M.5.5										
M.5.6										
M.5.7	14(1)	15(2)	25(2)	24(1)						
NBT.NF.M.5B										
M.5.8	28(2)	27(1)								
M.5.9	10(3)									
M.5.10	5(3)	29(1)	30(1)	24(1)	18(2)	17(1)	32(1)			
NBT.NF.M.5C										
M.5.11	31(2)	24(1)								
M.5.12	23(1)	30(1)	17(2)	16(1)	31(1)	6(6)				
NBT.NF.M.5D										
M.5.13	8(3)	21(2)	20(1)							
M.5.14	1(2)	11(3)	6(4)							
M.5.15	1(1)									

M.5.16	33(4)									
M.5.17	33(2)									
MD.G.M.5										
MD.G.M.5A										
M.5.18	12(2)	30(1)								
MD.G.M.5B										
M.5.19	3(2)									
MD.G.M.5C										
M.5.20										
M.5.21	23(2)	22(1)								
M.5.22	16(2)	15(1)	34(3)							
MD.G.M.5D										
M.5.23	3(1)	25(1)	26(2)							
M.5.24	2(3)									
MD.G.M.5E										
M.5.25	7(1)	20(2)	19(1)							
M.5.26	7(2)									
MHM										
MHM1										
MHM2										
MHM3										
MHM4										
MHM5										
MHM6										
MHM7										
MHM8										

Table 5.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 WV MATH 2019 Grade 5 B4

Low		Medium		High	
1.2		3.6		6	
1 10199-1604	M.5.14:2	M.5.15:1			
2 10199-1030	M.5.24:3				
3 10199-1103	M.5.19:2	M.5.23:1			
4 REP10199-1056	M.5.3:3				
5 10199-1683	M.5.10:3				
6 10199-12624	M.5.12:6	M.5.14:4			
7 10199-1072	M.5.25:1	M.5.26:2			
8 10199-1521	M.5.13:3				
9 10199-1109	M.5.2:3				
10 10199-1442	M.5.9:3				
11 10199-1650	M.5.14:3				
12 10199-1122	M.5.2:1	M.5.18:2			
13 10199-1078	M.5.1:1	M.5.2:1	M.5.3:1		
14 10199-1144	M.5.3:2	M.5.7:1			
15 10199-1142	M.5.7:2	M.5.22:1			
16 10199-2359	M.5.12:1	M.5.22:2			
17 10199-1515	M.5.10:1	M.5.12:2			
18 10199-1028	M.5.1:1	M.5.10:2			
19 10199-10129	M.5.1:1	M.5.2:1	M.5.25:1		
20 10199-1197	M.5.13:1	M.5.25:2			
21 REP10199-1712	M.5.2:1	M.5.13:2			
22 10199-1114	M.5.2:2	M.5.21:1			
23 REP10199-7468	M.5.12:1	M.5.21:2			
24 10199-1684	M.5.7:1	M.5.10:1	M.5.11:1		
25 10199-1155	M.5.7:2	M.5.23:1			
26 10199-1298	M.5.2:1	M.5.23:2			
27 10199-1680	M.5.1:1	M.5.2:1	M.5.8:1		
28 10199-1019	M.5.2:1	M.5.8:2			
29 10199-1137	M.5.2:2	M.5.10:1			
30 10199-1130	M.5.10:1	M.5.12:1	M.5.18:1		
31 10199-1517	M.5.11:2	M.5.12:1			
32 10199-999	M.5.4:2	M.5.10:1			
33 10199-1723	M.5.16:4	M.5.17:2			
34 10199-1902	M.5.22:3				

Table 5.9

Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])

WV MATH 2019 Grade 5 B4

Low DOK		Matched DOK		High DOK

OA.M.5										
OA.M.5A										
M.5.1: [1]	13:(1)[2]	18:(1)[1]	19:(1)[2]	27:(1)[1]						
M.5.2: [2]	9:(3)[1]	12:(1)[1]	13:(1)[2]	19:(1)[2]	21:(1)[1]	22:(2)[2]	26:(1)[1]	27:(1)[2]	28:(1)[1]	29:(2)[1]
OA.M.5B										
M.5.3: [2]	4:(3)[2]	13:(1)[2]	14:(2)[2]							
NBT.NF.M.5										
NBT.NF.M.5A										
M.5.4: [2]	32:(2)[2]									
M.5.5										
M.5.6										
M.5.7: [1]	14:(1)[1]	15:(2)[2]	24:(1)[1]	25:(2)[1]						
NBT.NF.M.5B										
M.5.8: [1]	27:(1)[2]	28:(2)[2]								
M.5.9: [2]	10:(3)[1]									
M.5.10: [2]	5:(3)[2]	17:(1)[2]	18:(2)[2]	24:(1)[2]	29:(1)[1]	30:(1)[1]	32:(1)[2]			
NBT.NF.M.5C										
M.5.11: [1]	24:(1)[2]	31:(2)[2]								
M.5.12: [2]	6:(6)[2]	16:(1)[1]	17:(2)[2]	23:(1)[2]	30:(1)[2]	31:(1)[2]				
NBT.NF.M.5D										
M.5.13: [2]	8:(3)[1]	20:(1)[1]	21:(2)[1]							
M.5.14: [2]	1:(2)[2]	6:(4)[2]	11:(3)[2]							
M.5.15: [2]	1:(1)[2]									
M.5.16: [2]	33:(4)[2]									
M.5.17: [2]	33:(2)[2]									
MD.G.M.5										
MD.G.M.5A										
M.5.18: [2]	12:(2)[2]	30:(1)[2]								

MD.G.M.5B										
M.5.19: [2]	3:(2)[1]									
MD.G.M.5C										
M.5.20										
M.5.21: [1]	22:(1)[1]	23:(2)[2]								
M.5.22: [2]	15:(1)[2]	16:(2)[2]	34:(3)[2]							
MD.G.M.5D										
M.5.23: [1]	3:(1)[1]	25:(1)[1]	26:(2)[1]							
M.5.24: [2]	2:(3)[2]									
MD.G.M.5E										
M.5.25: [2]	7:(1)[1]	19:(1)[1]	20:(2)[2]							
M.5.26: [1]	7:(2)[2]									
MHM										
MHM1										
MHM2										
MHM3										
MHM4										
MHM5										
MHM6										
MHM7										
MHM8										

Batch 1 West Virginia Math Grade 6

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B1_v3 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	1 2	6 5	54.55 45.45	16	0	YES
EE.M.6 Expressions and Equations	3	9	1 2	5 4	55.56 44.44	13.83	0.41	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	1 2	3 6	33.33 66.67	7.17	0.41	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0.5	0.84	NO
Total	18	37	1 2 3	14 17 6	38 46 16	37.5	0.84	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B1_v3 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	16	0	21.88	9	65.62	9	12.5	12	YES
EE.M.6 Expressions and Equations	3	9	13.83	0.41	4.76	6	85.44	7	9.8	8	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	7.17	0.41	23.21	7	67.86	15	8.93	11	YES
MHM Mathematical Habits of Mind	8	8	0.5	0.84	25	35	75	35	0	0	YES
Total	18	37	37.5	0.84	16	5.5	73.33	7.5	10.67	8.6	

Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B1_v3 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	16	0	10.33	0.82	93.94	7.42	YES	43	1	0.77	0.04	YES
EE.M.6 Expressions and Equations	3	9	13.83	0.41	6.83	0.98	75.93	10.92	YES	34	2	0.74	0.01	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	7.17	0.41	6	0	66.67	0	YES	21	1	0.87	0.02	YES
MHM Mathematical Habits of Mind	8	8	0.5	0.84	0.5	0.84	6.25	10.46	NO	1	2	0.92	0.2	YES
Total	18	37	37.5	0.84	5.9	4.07	60.7	38		25	18	0.82	0.09	

Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 6 B1_v3 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	YES	YES	YES	YES
EE.M.6 Expressions and Equations	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	YES	NO	YES

Table 6.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 6 B1_v3

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	1	1	1	1	1	1
2	1	1	1	1	1	2
3	2	2	2	2	2	2
4	2	1	1	2	1	2
5	1	2	2	1	1	2
6	2	2	2	2	2	2
7	2	2	2	1	1	1
8	1	1	1	1	1	1
9	2	2	2	2	2	2
10	1	1	1	2	1	1
11	1	1	1	1	1	1
12	1	1	1	1	1	2
13	2	2	2	2	2	2
14	1	1	1	1	1	1
15	1	2	1	1	1	1
16	2	2	2	2	2	2
17	1	2	2	2	1	2
18	2	2	2	1	1	2
19	2	2	2	2	2	2
20	1	1	1	1	1	1
21	2	2	2	2	2	2
22	2	2	1	2	1	1
23	1	1	1	1	1	2
24	1	1	1	1	1	1
25	2	2	2	2	2	2
26	1	1	1	2	1	2
27	1	1	2	1	1	2
28	2	2	2	2	2	3
29	1	1	1	1	1	1
30	1	1	1	1	1	2
31	1	1	1	1	1	1
32	1	1	1	1	1	1
33	2	2	2	2	2	2
34	1	1	1	1	1	1

Intraclass correlation - .9276

Pairwise Comparison - 0.8

Table 6.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 6 B1_v3 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	M.6.22			1	M.6.22			1	M.6.22			1	M.6.22			1	M.6.22			1	M.6.22		
2	1	M.6.26			1	M.6.26			1	M.6.26			1	M.6.26			2	M.6.26						
3	2	M.6.24			2	M.6.24			2	M.6.24			2	M.6.24			2	M.6.24			2	M.6.24		
4	2	M.6.28			1	M.6.28			1	M.6.28			2	M.6.28			1	M.6.28			2	M.6.28		
5	1	M.6.21			2	M.6.21			2	M.6.21			1	M.6.21			1	M.6.21			2	M.6.21		
6	2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25		
7	2	M.6.17			2	M.6.17			2	M.6.17			1	M.6.17			1	M.6.17			1	M.6.21		
8	1	M.6.26			1	M.6.26			1	M.6.26			1	M.6.26			1	M.6.26			1	M.6.26		
9	2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18		
10	1	M.6.14			1	M.6.14			1	M.6.14			2	M.6.14			1	M.6.14			1	M.6.14		
11	1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1		
12	1	M.6.3			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			2	M.6.2		
13	2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.20			2	M.6.18			2	M.6.20		
14	1	M.6.4			1	M.6.4			1	M.6.4			1	M.6.4			1	M.6.4			1	M.6.4		
15	1	M.6.1			2	M.6.1			1	M.6.1			1	M.6.3			1	M.6.1			1	M.6.1		
16	2	M.6.16			2	M.6.16			2	M.6.16			2	M.6.19			2	M.6.19			2	M.6.19		
17	1	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			1	M.6.18			2	M.6.16		
18	2	M.6.9			2	M.6.9			2	M.6.9			1	M.6.11			1	M.6.9			2	M.6.9		
19	2	M.6.17			2	M.6.17			2	M.6.17			2	M.6.17			2	M.6.17			2	M.6.17		
20	1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5		
21	2	M.6.3			2	M.6.3	MHM6		2	M.6.3			2	M.6.3			2	M.6.3			2	M.6.3		
22	2	M.6.11			2	M.6.11			1	M.6.11			2	M.6.11			1	M.6.9			1	M.6.11		
23	1	M.6.15			1	M.6.14			1	M.6.14			1	M.6.15			1	M.6.14			2	M.6.14		
24	1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7		
25	2	M.6.3			2	M.6.3			2	M.6.3			2	M.6.3			2	M.6.3			2	M.6.2		
26	1	M.6.12			1	M.6.12			1	M.6.12			2	M.6.12			1	M.6.12			2	M.6.12		
27	1	M.6.10			1	M.6.10			2	M.6.10			1	M.6.10			1	M.6.10			2	M.6.10		
28	2	M.6.3			2	M.6.3	MHM1		2	M.6.3			2	M.6.3			2	M.6.3			3	M.6.3	MHM1	
29	1	M.6.13			1	M.6.12			1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.12		
30	1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.3			2	M.6.3		
31	1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13		
32	1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6		
33	2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18		
34	1	M.6.8			1	M.6.8			1	M.6.8			1	M.6.8			1	M.6.9			1	M.6.9		
Objective Pairwise Comparison: 0.82																								
Standard Pairwise Comparison: 0.97																								

Table 6.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV MATH 2019 Grade 6 B1_v3

Low		Medium		High	
0		7.2		12	
RP.NS.M.6					
RP.NS.M.6A					
M.6.1	11(6)	15(5)			
M.6.2	12(5)	25(1)	30(4)		
M.6.3	30(2)	28(6)	25(5)	21(6)	12(1)
RP.NS.M.6B					
M.6.4	14(6)				
RP.NS.M.6C					
M.6.5	20(6)				
M.6.6	32(6)				
M.6.7	24(6)				
RP.NS.M.6D					
M.6.8	34(4)				
M.6.9	34(2)	18(10)	22(1)		
M.6.10	27(6)				
M.6.11	22(5)	18(2)			
EE.M.6					
EE.M.6A					
M.6.12	26(6)	29(2)			
M.6.13	29(4)	31(6)			
M.6.14	23(4)	10(6)			
M.6.15	23(2)				
EE.M.6B					

M.6.16	17(1)	16(3)				
M.6.17	7(5)	19(6)				
M.6.18	17(5)	9(12)	13(4)	33(12)		
M.6.19	16(3)					
EE.M.6C						
M.6.20	13(2)					
G.SP.M.6						
G.SP.M.6.A						
M.6.21	7(1)	5(6)				
M.6.22	1(6)					
M.6.23						
M.6.24	3(6)					
G.SP.M.6.B						
M.6.25	6(6)					
M.6.26	8(6)	2(6)				
M.6.27						
G.SP.M.6.C						
M.6.28	4(6)					
M.6.29						
MHM						
MHM1	28(2)					
MHM2						
MHM3						
MHM4						
MHM5						
MHM6	21(1)					
MHM7						
MHM8						

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

WV MATH 2019 Grade 6 B1_v3

Low		Medium		High	
2.4		7.2		12	
1	10199-9773	M.6.22:6			
2	10199-9809	M.6.26:6			
3	10199-10846	M.6.24:6			
4	REP10199-2012	M.6.28:6			
5	10199-1997	M.6.21:6			
6	REP10199-2033	M.6.25:6			
7	REP10199-2008	M.6.17:5		M.6.21:1	
8	10199-2415	M.6.26:6			
9	10199-9748	M.6.18:12			
10	10199-1968	M.6.14:6			
11	10199-9956	M.6.1:6			
12	10199-1890	M.6.2:5		M.6.3:1	
13	10199-1982	M.6.18:4		M.6.20:2	
14	10199-1462	M.6.4:6			
15	10199-1577	M.6.1:5		M.6.3:1	
16	10199-2453	M.6.16:3		M.6.19:3	
17	REP10199-1854	M.6.16:1		M.6.18:5	
18	10199-10848	M.6.9:10		M.6.11:2	
19	10199-2070	M.6.17:6			
20	10199-1537	M.6.5:6			
21	REP10199-10156	M.6.3:6		MHM6:1	
22	10199-13341	M.6.9:1		M.6.11:5	
23	10199-9887	M.6.14:4		M.6.15:2	
24	10199-1559	M.6.7:6			
25	10199-10150	M.6.2:1		M.6.3:5	
26	REP10199-1956	M.6.12:6			
27	10199-2093	M.6.10:6			
28	10199-1784	M.6.3:6		MHM1:2	
29	10199-2028	M.6.12:2		M.6.13:4	
30	10199-9973	M.6.2:4		M.6.3:2	
31	10199-1914	M.6.13:6			
32	10199-1739	M.6.6:6			
33	10199-1892	M.6.18:12			
34	10199-1707	M.6.8:4		M.6.9:2	

Table 6.9

Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
 WV MATH 2019 Grade 6 B1_v3

Low DOK		Matched DOK		High DOK

RP.NS.M.6						
RP.NS.M.6A						
M.6.1: [1]	11:(6)[1]	15:(5)[1]				
M.6.2: [1]	12:(5)[1]	25:(1)[2]	30:(4)[1]			
M.6.3: [2]	12:(1)[1]	15:(1)[1]	21:(6)[2]	25:(5)[2]	28:(6)[2]	30:(2)[2]
RP.NS.M.6B						
M.6.4: [2]	14:(6)[1]					
RP.NS.M.6C						
M.6.5: [1]	20:(6)[1]					
M.6.6: [1]	32:(6)[1]					
M.6.7: [1]	24:(6)[1]					
RP.NS.M.6D						
M.6.8: [2]	34:(4)[1]					
M.6.9: [1]	18:(10)[2]	22:(1)[1]	34:(2)[1]			
M.6.10: [2]	27:(6)[1]					
M.6.11: [2]	18:(2)[1]	22:(5)[2]				
EE.M.6						
EE.M.6A						
M.6.12: [1]	26:(6)[1]	29:(2)[1]				
M.6.13: [1]	29:(4)[1]	31:(6)[1]				
M.6.14: [1]	10:(6)[1]	23:(4)[1]				
M.6.15: [1]	23:(2)[1]					
EE.M.6B						
M.6.16: [1]	16:(3)[2]	17:(1)[2]				
M.6.17: [2]	7:(5)[2]	19:(6)[2]				
M.6.18: [2]	9:(12)[2]	13:(4)[2]	17:(5)[2]	33:(12)[2]		
M.6.19: [2]	16:(3)[2]					
EE.M.6C						
M.6.20: [2]	13:(2)[2]					
G.SP.M.6						
G.SP.M.6.A						
M.6.21: [2]	5:(6)[2]	7:(1)[1]				
M.6.22: [2]	1:(6)[1]					
M.6.23						
M.6.24: [2]	3:(6)[2]					
G.SP.M.6.B						
M.6.25: [2]	6:(6)[2]					
M.6.26: [1]	2:(6)[1]	8:(6)[1]				
M.6.27						
G.SP.M.6.C						
M.6.28: [1]	4:(6)[2]					
M.6.29						
MHM						

MHM1: [3]	28:(2)[2]					
MHM2						
MHM3						
MHM4						
MHM5						
MHM6: [2]	21:(1)[2]					
MHM7						
MHM8						

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 6B1 MATH v2

[illegible]

Table 6.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 6B1 MATH v2

Low		Medium		High	
2.4		7.2		12	
1	10199-9773	Exact:6			
2	10199-9809	Exact:1		Partial:5	
3	10199-10846	Exact:6			
4	REP10199-2012	Exact:6			
5	10199-1997	Exact:6			
6	REP10199-2033	Exact:6			
7	REP10199-2008	Exact:1		Partial:5	
8	10199-2415	Exact:5		Minimal:1	
9	10199-9748	Exact:12			
10	10199-1968	Exact:6			
11	10199-9956	Exact:6			
12	10199-1890	Exact:6			
13	10199-1982	Exact:3		Partial:2	
14	10199-1462	Exact:6		Minimal:1	
15	10199-1577	Exact:5		Partial:1	
16	10199-2453	Exact:3		Partial:3	
17	REP10199-1854	Partial:4		Minimal:2	
18	10199-10848	Exact:10		Partial:2	
19	10199-2070	Exact:6			
20	10199-1537	Exact:6			
21	REP10199-10156	Exact:5			
22	10199-13341	Exact:6			
23	10199-9887	Exact:4		Partial:2	
24	10199-1559	Exact:6			
25	10199-10150	Exact:6			
26	REP10199-1956	Exact:6			
27	10199-2093	Exact:6			
28	10199-1784	Exact:6			
29	10199-2028	Exact:4		Partial:2	
30	10199-9973	Exact:6			
31	10199-1914	Exact:6			
32	10199-1739	Exact:6			
33	10199-1892	Exact:12			
34	10199-1707	Exact:3		Partial:3	

Batch 2 West Virginia Math Grade 6 – Elementary Panel

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B2 ELEM GROUP Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	1 2	6 5	54.55 45.45	15.17	1.17	YES
EE.M.6 Expressions and Equations	3	9	1 2	5 4	55.56 44.44	11.67	0.82	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	1 2	3 6	33.33 66.67	7.83	0.41	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0.33	0.52	NO
Total	18	37	1 2 3	14 17 6	38 46 16	35	0.89	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B2 ELEM GROUP Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	15.17	1.17	22.74	10	65.11	8	12.15	5	YES
EE.M.6 Expressions and Equations	3	9	11.67	0.82	16.72	13	80.48	12	2.8	4	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	7.83	0.41	8.33	6	91.67	6	0	0	YES
MHM Mathematical Habits of Mind	8	8	0.33	0.52	0	0	100	0	0	0	YES
Total	18	37	35	0.89	17.62	8.1	76.19	6	6.19	3.4	

Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B2 ELEM GROUP Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	15.17	1.17	8.83	0.41	80.3	3.71	YES	43	3	0.8	0.01	YES
EE.M.6 Expressions and Equations	3	9	11.67	0.82	8	0	88.89	0	YES	33	3	0.81	0.02	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	7.83	0.41	4.83	0.41	53.7	4.54	YES	22	1	0.8	0.03	YES
MHM Mathematical Habits of Mind	8	8	0.33	0.52	0.33	0.52	4.17	6.45	NO	1	1	1	0	YES
Total	18	37	35	0.89	5.5	3.85	56.76	38		25	18	0.85	0.1	

Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 6 B2 ELEM GROUP Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	YES	YES	YES	YES
EE.M.6 Expressions and Equations	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	YES	NO	YES

Table 6.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 6 B2 ELEM GROUP

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	2	2	1	2	2
2	1	1	1	1	1	1
3	2	2	1	2	2	2
4	2	2	2	2	2	2
5	2	2	2	2	2	2
6	1	1	1	1	1	1
7	2	2	2	2	2	2
8	1	2	2	2	2	1
9	1	1	1	1	1	1
10	1	1	1	1	1	1
11	2	2	1	1	1	1
12	1	1	1	2	1	1
13	1	1	1	2	2	2
14	1	1	1	1	1	1
15	2	2	2	1	2	1
16	1	2	2	2	2	2
17	1	2	1	1	2	2
18	1	2	2	1	1	1
19	1	1	1	1	1	1
20	1	1	1	1	1	1
21	1	1	1	2	1	2
22	2	2	2	1	2	2
23	2	2	2	2	2	2
24	1	2	1	1	1	1
25	1	1	1	1	1	1
26	1	2	1	1	1	1
27	1	1	1	1	1	1
28	1	2	1	1	1	1
29	1	1	1	1	1	1
30	1	1	1	1	1	1
31	1	1	1	1	1	1
32	3	3	3	3	3	2
33	1	1	1	1	1	1
34	1	1	2	1	1	1

Intraclass correlation - .9299

Pairwise Comparison - 0.79

Table 6.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 6 B2 ELEM GROUP Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.6.21			2	M.6.21			2	M.6.21			1	M.6.21			2	M.6.21			2	M.6.21		
2	1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28		
3	2	M.6.22			2	M.6.22			1	M.6.22			2	M.6.22			2	M.6.22			2	M.6.22		
4	2	M.6.21	M.6.17		2	M.6.21			2	M.6.21			2	M.6.21	M.6.17		2	M.6.21			2	M.6.17	M.6.21	
5	2	M.6.25			2	M.6.3			2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25		
6	1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28		
7	2	M.6.21			2	M.6.21			2	M.6.21			2	M.6.21			2	M.6.23			2	M.6.23		
8	1	M.6.23			2	M.6.23			2	M.6.23			2	M.6.23			2	M.6.23			1	M.6.23		
9	1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.5			1	M.6.2		
10	1	M.6.19			1	M.6.14			1	M.6.14			1	M.6.19			1	M.6.19			1	M.6.19		
11	2	M.6.3			2	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15		
12	1	M.6.5			1	M.6.1			1	M.6.1			2	M.6.3			1	M.6.1			1	M.6.1		
13	1	M.6.3			1	M.6.3			1	M.6.3			2	M.6.3			2	M.6.5			2	M.6.3		
14	1	M.6.16			1	M.6.16			1	M.6.16			1	M.6.16			1	M.6.18			1	M.6.16		
15	2	M.6.4			2	M.6.4			2	M.6.4			1	M.6.4			2	M.6.20			1	M.6.4		
16	1	M.6.20			2	M.6.20			2	M.6.20			2	M.6.20			2	M.6.13			2	M.6.20		
17	1	M.6.11			2	M.6.9			1	M.6.11			1	M.6.11			2	M.6.11			2	M.6.11		
18	1	M.6.11			2	M.6.18			2	M.6.18			1	M.6.18			1	M.6.17			1	M.6.18		
19	1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15		
20	1	M.6.1			1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5		
21	1	M.6.1			1	M.6.1			1	M.6.1			2	M.6.1			1	M.6.1			2	M.6.1		
22	2	M.6.17			2	M.6.17			2	M.6.17			1	M.6.17			2	M.6.17			2	M.6.17		
23	2	M.6.10			2	M.6.10			2	M.6.10			2	M.6.10			2	M.6.10			2	M.6.10		
24	1	M.6.12			2	M.6.4			1	M.6.4			1	M.6.4			1	M.6.12			1	M.6.4		
25	1	M.6.15			1	M.6.12			1	M.6.12			1	M.6.12			1	M.6.12			1	M.6.12		
26	1	M.6.7			2	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7		
27	1	M.6.2			1	M.6.1			1	M.6.1			1	M.6.2			1	M.6.2			1	M.6.2		
28	1	M.6.18			2	M.6.18			1	M.6.16			1	M.6.16			1	M.6.16			1	M.6.18		
29	1	M.6.10			1	M.6.10			1	M.6.10			1	M.6.10			1	M.6.10			1	M.6.10		
30	1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6		
31	1	M.6.12			1	M.6.12			1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.12		
32	3	M.6.3	MHM1		3	M.6.3	MHM1		3	M.6.3			3	M.6.3			3	M.6.3			2	M.6.3	M.6.1	
33	1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13		
34	1	M.6.3			1	M.6.6			2	M.6.3			1	M.6.3			1	M.6.6			1	M.6.3		
Objective Pairwise Comparison: 0.72																								
Standard Pairwise Comparison: 0.92																								

Table 6.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV MATH 2019 Grade 6 B2 ELEM GROUP

Low		Medium		High	
0		3.6		6	
RP.NS.M.6					
RP.NS.M.6A					
M.6.1	12(4)	20(1)	21(6)	27(2)	32(1)
M.6.2	27(4)	9(5)			
M.6.3	12(1)	11(1)	13(5)	5(1)	32(6)
RP.NS.M.6B					
M.6.4	24(4)	15(5)			
RP.NS.M.6C					
M.6.5	13(1)	12(1)	9(1)	20(5)	
M.6.6	30(6)	34(2)			
M.6.7	26(6)				
RP.NS.M.6D					
M.6.8					
M.6.9	17(1)				
M.6.10	23(6)	29(6)			
M.6.11	17(5)	18(1)			
EE.M.6					
EE.M.6A					
M.6.12	24(2)	25(5)	31(3)		
M.6.13	33(6)	16(1)			
M.6.14	10(2)				
M.6.15	11(5)	31(3)	25(1)	19(6)	
EE.M.6B					

M.6.16	28(3)	14(5)				
M.6.17	18(1)	22(6)	4(3)			
M.6.18	18(4)	28(3)	14(1)			
M.6.19	10(4)					
EE.M.6C						
M.6.20	16(5)	15(1)				
G.SP.M.6						
G.SP.M.6.A						
M.6.21	1(6)	7(4)	4(6)			
M.6.22	3(6)					
M.6.23	7(2)	8(6)				
M.6.24						
G.SP.M.6.B						
M.6.25	5(5)					
M.6.26						
M.6.27						
G.SP.M.6.C						
M.6.28	6(6)	2(6)				
M.6.29						
MHM						
MHM1	32(2)					
MHM2						
MHM3						
MHM4						
MHM5						
MHM6						
MHM7						
MHM8						

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 WV MATH 2019 Grade 6 B2 ELEM GROUP

Low		Medium		High	
1.2		3.6		6	
1	10199-1998	M.6.21:6			
2	10199-2306	M.6.28:6			
3	10199-1868	M.6.22:6			
4	REP10199-2013	M.6.17:3	M.6.21:6		
5	REP10199-2033	M.6.3:1	M.6.25:5		
6	REP10199-2012	M.6.28:6			
7	10199-2191	M.6.21:4	M.6.23:2		
8	10199-2145	M.6.23:6			
9	10199-1923	M.6.2:5	M.6.5:1		
10	10199-1785	M.6.14:2	M.6.19:4		
11	10199-1788	M.6.3:1	M.6.15:5		
12	10199-2078	M.6.1:4	M.6.3:1	M.6.5:1	
13	REP10199-10156	M.6.3:5	M.6.5:1		
14	10199-13345	M.6.16:5	M.6.18:1		
15	REP10199-10208	M.6.4:5	M.6.20:1		
16	10199-1980	M.6.13:1	M.6.20:5		
17	10199-1836	M.6.9:1	M.6.11:5		
18	10199-1887	M.6.11:1	M.6.17:1	M.6.18:4	
19	10199-2106	M.6.15:6			
20	10199-1546	M.6.1:1	M.6.5:5		
21	10199-1568	M.6.1:6			
22	10199-1872	M.6.17:6			
23	10199-2130	M.6.10:6			
24	10199-1463	M.6.4:4	M.6.12:2		
25	10199-1938	M.6.12:5	M.6.15:1		
26	10199-1962	M.6.7:6			
27	10199-1941	M.6.1:2	M.6.2:4		
28	REP10199-1854	M.6.16:3	M.6.18:3		
29	10199-2125	M.6.10:6			
30	10199-1589	M.6.6:6			
31	REP10199-1956	M.6.12:3	M.6.15:3		
32	10199-2004	M.6.1:1	M.6.3:6	M.6.5:1	
33	10199-2118	M.6.13:6			
34	10199-1819	M.6.3:4	M.6.6:2		

Table 6.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV MATH 2019 Grade 6 B2 ELEM GROUP

Low DOK		Matched DOK		High DOK

RP.NS.M.6						
RP.NS.M.6A						
M.6.1: [1]	12:(4)[1]	20:(1)[1]	21:(6)[1]	27:(2)[1]	32:(1)[2]	
M.6.2: [1]	9:(5)[1]	27:(4)[1]				
M.6.3: [2]	5:(1)[2]	11:(1)[2]	12:(1)[2]	13:(5)[1]	32:(6)[3]	34:(4)[1]
RP.NS.M.6B						
M.6.4: [2]	15:(5)[2]	24:(4)[1]				
RP.NS.M.6C						
M.6.5: [1]	9:(1)[1]	12:(1)[1]	13:(1)[2]	20:(5)[1]		
M.6.6: [1]	30:(6)[1]	34:(2)[1]				
M.6.7: [1]	26:(6)[1]					
RP.NS.M.6D						
M.6.8						
M.6.9: [1]	17:(1)[2]					
M.6.10: [2]	23:(6)[2]	29:(6)[1]				
M.6.11: [2]	17:(5)[1]	18:(1)[1]				
EE.M.6						
EE.M.6A						
M.6.12: [1]	24:(2)[1]	25:(5)[1]	31:(3)[1]			
M.6.13: [1]	16:(1)[2]	33:(6)[1]				
M.6.14: [1]	10:(2)[1]					
M.6.15: [1]	11:(5)[1]	19:(6)[1]	25:(1)[1]	31:(3)[1]		
EE.M.6B						
M.6.16: [1]	14:(5)[1]	28:(3)[1]				
M.6.17: [2]	4:(3)[2]	18:(1)[1]	22:(6)[2]			
M.6.18: [2]	14:(1)[1]	18:(4)[2]	28:(3)[1]			
M.6.19: [2]	10:(4)[1]					
EE.M.6C						

M.6.20: [2]	15:(1)[2]	16:(5)[2]				
G.SP.M.6						
G.SP.M.6.A						
M.6.21: [2]	1:(6)[2]	4:(6)[2]	7:(4)[2]			
M.6.22: [2]	3:(6)[2]					
M.6.23: [2]	7:(2)[2]	8:(6)[2]				
M.6.24						
G.SP.M.6.B						
M.6.25: [2]	5:(5)[2]					
M.6.26						
M.6.27						
G.SP.M.6.C						
M.6.28: [1]	2:(6)[1]	6:(6)[1]				
M.6.29						
MHM						
MHM1: [3]	32:(2)[3]					
MHM2						
MHM3						
MHM4						
MHM5						
MHM6						
MHM7						
MHM8						

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 6B2 MATH ELEM GROUP

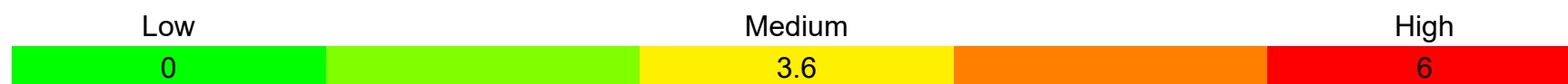
[illegible]

Table 6.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 6B2 MATH ELEM GROUP

Low		Medium		High	
1.2		3.6		6	
1	10199-1998	Exact:5			
2	10199-2306	Negligible:6			
3	10199-1868	Exact:3		Partial:3	
4	REP10199-2013	Exact:5		Partial:1	
5	REP10199-2033	Exact:6			
6	REP10199-2012	Exact:6			
7	10199-2191	Exact:2		Partial:4	
8	10199-2145	Exact:5		Partial:1	
9	10199-1923	Exact:5		Partial:1	
10	10199-1785	Exact:5		Partial:1	
11	10199-1788	Partial:5		Minimal:1	
12	10199-2078	Exact:4		Partial:2	
13	REP10199-10156	Exact:5		Partial:1	
14	10199-13345	Exact:6			
15	REP10199-10208	Exact:5		Partial:1	
16	10199-1980	Exact:6			
17	10199-1836	Exact:6			
18	10199-1887	Exact:3		Partial:3	
19	10199-2106	Partial:5		Minimal:1	
20	10199-1546	Exact:6			
21	10199-1568	Exact:6			
22	10199-1872	Exact:6			
23	10199-2130	Exact:5		Partial:1	
24	10199-1463	Exact:6			
25	10199-1938	Exact:5		Partial:1	
26	10199-1962	Exact:4		Partial:2	
27	10199-1941	Exact:4		Partial:2	
28	REP10199-1854	Partial:1		Minimal:5	
29	10199-2125	Exact:3		Partial:2	
30	10199-1589	Exact:6		Minimal:1	
31	REP10199-1956	Exact:3		Partial:3	
32	10199-2004	Exact:6			
33	10199-2118	Exact:6			
34	10199-1819	Exact:4		Partial:2	

Batch 2 West Virginia Math Grade 6 – Middle Grades Panel

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B2_v2 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	1 2	6 5	54.55 45.45	15.33	0.82	YES
EE.M.6 Expressions and Equations	3	9	1 2	5 4	55.56 44.44	11.83	0.75	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	1 2	3 6	33.33 66.67	8	0.63	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	2.67	1.21	NO
Total	18	37	1 2 3	14 17 6	38 46 16	37.83	1.17	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B2_v2 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	15.33	0.82	15.16	5	73.86	10	10.98	8	YES
EE.M.6 Expressions and Equations	3	9	11.83	0.75	9.63	6	79.56	14	10.81	12	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	8	0.63	3.94	6	90.05	12	6.02	7	YES
MHM Mathematical Habits of Mind	8	8	2.67	1.21	36.11	29	63.89	29	0	0	YES
Total	18	37	37.83	1.17	13.22	3.2	77.53	9.1	9.25	7.3	

Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B2_v2 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	15.33	0.82	9.17	0.41	83.33	3.71	YES	41	3	0.81	0.03	YES
EE.M.6 Expressions and Equations	3	9	11.83	0.75	8.33	0.52	92.59	5.74	YES	31	2	0.83	0.02	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	8	0.63	5	0.63	55.56	7.03	YES	21	1	0.77	0.01	YES
MHM Mathematical Habits of Mind	8	8	2.67	1.21	2.5	1.05	31.25	13.11	NO	7	3	0.94	0.09	YES
Total	18	37	37.83	1.17	6.2	3.08	65.68	28		25	14	0.84	0.07	

Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 6 B2_v2 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	YES	YES	YES	YES
EE.M.6 Expressions and Equations	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	YES	NO	YES

Table 6.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 6 B2_v2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	2	2	2	2	1
2	1	1	1	1	1	1
3	2	2	2	1	2	2
4	2	2	2	2	2	2
5	2	2	2	2	2	2
6	2	1	1	2	1	2
7	2	2	2	2	2	2
8	2	2	2	2	2	2
9	1	1	1	1	1	1
10	1	1	1	1	2	1
11	1	2	1	1	1	2
12	2	1	1	1	1	1
13	2	2	2	2	2	2
14	1	1	1	2	2	1
15	1	2	1	1	1	1
16	2	2	2	2	2	1
17	1	1	1	1	1	2
18	2	2	2	2	2	2
19	1	1	1	1	1	1
20	1	1	1	1	1	1
21	1	1	1	1	1	2
22	2	2	2	2	2	2
23	2	2	2	2	2	2
24	1	1	1	1	1	1
25	1	1	1	1	1	1
26	1	1	1	1	1	1
27	1	1	1	1	1	1
28	1	2	2	2	1	2
29	2	2	2	2	2	2
30	1	1	1	1	1	1
31	1	1	1	2	1	2
32	3	3	3	3	3	3
33	1	1	1	1	1	1
34	2	2	1	2	1	2

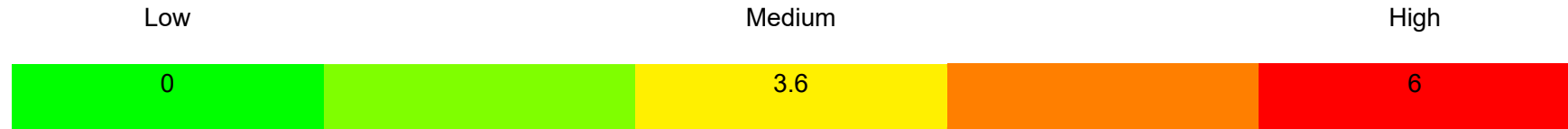
Intraclass correlation - .9498

Pairwise Comparison - 0.83

Table 6.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 6 B2_v2 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.6.21			2	M.6.21			2	M.6.21			2	M.6.21			2	M.6.21			1	M.6.21		
2	1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28		
3	2	M.6.22			2	M.6.22	MHM4		2	M.6.22	MHM4		1	M.6.22	MHM4		2	M.6.22			2	M.6.22	MHM4	
4	2	M.6.21	M.6.17		2	M.6.17	M.6.21		2	M.6.21			2	M.6.17	M.6.21		2	M.6.17	M.6.21		2	M.6.13	M.6.21	
5	2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25	M.6.29	
6	2	M.6.28			1	M.6.28			1	M.6.28			2	M.6.28			1	M.6.28			2	M.6.28		
7	2	M.6.21	MHM6		2	M.6.21			2	M.6.21			2	M.6.21	MHM5		2	M.6.21			2	M.6.21		
8	2	M.6.23			2	M.6.23	MHM5		2	M.6.23			2	M.6.23			2	M.6.11			2	M.6.23		
9	1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2		
10	1	M.6.19			1	M.6.19			1	M.6.19			1	M.6.19			2	M.6.19			1	M.6.19		
11	1	M.6.15			2	M.6.15			1	M.6.15	MHM1		1	M.6.14			1	M.6.14			2	M.6.14	M.6.15	
12	2	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1		
13	2	M.6.3			2	M.6.3	MHM6		2	M.6.3	MHM6		2	M.6.3			2	M.6.3			2	M.6.2		
14	1	M.6.16			1	M.6.16			1	M.6.16			2	M.6.12			2	M.6.16			1	M.6.16		
15	1	M.6.4			2	M.6.4			1	M.6.3			1	M.6.4			1	M.6.4			1	M.6.4		
16	2	M.6.20			2	M.6.20			2	M.6.20			2	M.6.20			2	M.6.20			1	M.6.20		
17	1	M.6.11			1	M.6.9			1	M.6.9			1	M.6.9			1	M.6.9			2	M.6.9		
18	2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18		
19	1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15		
20	1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5		
21	1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.2			1	M.6.1			2	M.6.1		
22	2	M.6.17			2	M.6.17			2	M.6.17			2	M.6.17			2	M.6.17			2	M.6.17		
23	2	M.6.10			2	M.6.10			2	M.6.10			2	M.6.10			2	M.6.10			2	M.6.10		
24	1	M.6.4			1	M.6.4			1	M.6.4			1	M.6.4			1	M.6.4			1	M.6.4		
25	1	M.6.12			1	M.6.12			1	M.6.12			1	M.6.12			1	M.6.12			1	M.6.12		
26	1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7		
27	1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2		
28	1	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			1	M.6.7			2	M.6.18		
29	2	M.6.10			2	M.6.10			2	M.6.10			2	M.6.10			2	M.6.10			2	M.6.10		
30	1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6		
31	1	M.6.15			1	M.6.12			1	M.6.12			2	M.6.15			1	M.6.12			2	M.6.12		
32	3	M.6.3	MHM1		3	M.6.3	MHM1		3	M.6.3	MHM1		3	M.6.3	MHM1		3	M.6.3	MHM1		3	M.6.3	MHM1	
33	1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13		
34	2	M.6.3			2	M.6.3			1	M.6.3			2	M.6.3			1	M.6.3			2	M.6.3		
Objective Pairwise Comparison: 0.82																								
Standard Pairwise Comparison: 0.9																								

Table 6.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV MATH 2019 Grade 6 B2_v2



RP.NS.M.6				
RP.NS.M.6A				
M.6.1	12(6)	21(5)		
M.6.2	21(1)	27(6)	13(1)	9(6)
M.6.3	13(5)	15(1)	32(6)	34(6)
RP.NS.M.6B				
M.6.4	15(5)	24(6)		
RP.NS.M.6C				
M.6.5	20(6)			
M.6.6	30(6)			
M.6.7	28(1)	26(6)		
RP.NS.M.6D				
M.6.8				
M.6.9	17(5)			
M.6.10	23(6)	29(6)		
M.6.11	17(1)	8(1)		
EE.M.6				
EE.M.6A				
M.6.12	14(1)	25(6)	31(4)	
M.6.13	4(1)	33(6)		
M.6.14	11(3)			
M.6.15	31(2)	19(6)	11(4)	

EE.M.6B				
M.6.16	14(5)			
M.6.17	22(6)	4(4)		
M.6.18	18(6)	28(5)		
M.6.19	10(6)			
EE.M.6C				
M.6.20	16(6)			
G.SP.M.6				
G.SP.M.6.A				
M.6.21	1(6)	7(6)	4(6)	
M.6.22	3(6)			
M.6.23	8(5)			
M.6.24				
G.SP.M.6.B				
M.6.25	5(6)			
M.6.26				
M.6.27				
G.SP.M.6.C				
M.6.28	6(6)	2(6)		
M.6.29	5(1)			
MHM				
MHM1	32(6)	11(1)		
MHM2				
MHM3				
MHM4	3(4)			
MHM5	7(1)	8(1)		
MHM6	7(1)	13(2)		
MHM7				
MHM8				

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

WV MATH 2019 Grade 6 B2_v2

Low		Medium		High	
1.2		3.6		6	
1	10199-1998	M.6.21:6			
2	10199-2306	M.6.28:6			
3	10199-1868	M.6.22:6	MHM4:4		
4	REP10199-2013	M.6.13:1	M.6.17:4	M.6.21:6	
5	REP10199-2033	M.6.25:6	M.6.29:1		
6	REP10199-2012	M.6.28:6			
7	10199-2191	M.6.21:6	MHM5:1	MHM6:1	
8	10199-2145	M.6.11:1	M.6.23:5	MHM5:1	
9	10199-1923	M.6.2:6			
10	10199-1785	M.6.19:6			
11	10199-1788	M.6.14:3	M.6.15:4	MHM1:1	
12	10199-2078	M.6.1:6			
13	REP10199-10156	M.6.2:1	M.6.3:5	MHM6:2	
14	10199-13345	M.6.12:1	M.6.16:5		
15	REP10199-10208	M.6.3:1	M.6.4:5		
16	10199-1980	M.6.20:6			
17	10199-1836	M.6.9:5	M.6.11:1		
18	10199-1887	M.6.18:6			
19	10199-2106	M.6.15:6			
20	10199-1546	M.6.5:6			
21	10199-1568	M.6.1:5	M.6.2:1		
22	10199-1872	M.6.17:6			
23	10199-2130	M.6.10:6			
24	10199-1463	M.6.4:6			
25	10199-1938	M.6.12:6			
26	10199-1962	M.6.7:6			
27	10199-1941	M.6.2:6			
28	REP10199-1854	M.6.7:1	M.6.18:5		
29	10199-2125	M.6.10:6			
30	10199-1589	M.6.6:6			
31	REP10199-1956	M.6.12:4	M.6.15:2		
32	10199-2004	M.6.3:6	MHM1:6		
33	10199-2118	M.6.13:6			
34	10199-1819	M.6.3:6			

Table 6.9

Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])

WV MATH 2019 Grade 6 B2_v2

Low DOK		Matched DOK		High DOK

RP.NS.M.6				
RP.NS.M.6A				
M.6.1: [1]	12:(6)[1]	21:(5)[1]		
M.6.2: [1]	9:(6)[1]	13:(1)[2]	21:(1)[1]	27:(6)[1]
M.6.3: [2]	13:(5)[2]	15:(1)[1]	32:(6)[3]	34:(6)[2]
RP.NS.M.6B				
M.6.4: [2]	15:(5)[1]	24:(6)[1]		
RP.NS.M.6C				
M.6.5: [1]	20:(6)[1]			
M.6.6: [1]	30:(6)[1]			
M.6.7: [1]	26:(6)[1]	28:(1)[1]		
RP.NS.M.6D				
M.6.8				
M.6.9: [1]	17:(5)[1]			
M.6.10: [2]	23:(6)[2]	29:(6)[2]		
M.6.11: [2]	8:(1)[2]	17:(1)[1]		
EE.M.6				
EE.M.6A				
M.6.12: [1]	14:(1)[2]	25:(6)[1]	31:(4)[1]	
M.6.13: [1]	4:(1)[2]	33:(6)[1]		
M.6.14: [1]	11:(3)[1]			
M.6.15: [1]	11:(4)[2]	19:(6)[1]	31:(2)[2]	
EE.M.6B				
M.6.16: [1]	14:(5)[1]			
M.6.17: [2]	4:(4)[2]	22:(6)[2]		
M.6.18: [2]	18:(6)[2]	28:(5)[2]		
M.6.19: [2]	10:(6)[1]			
EE.M.6C				

M.6.20: [2]	16:(6)[2]			
G.SP.M.6				
G.SP.M.6.A				
M.6.21: [2]	1:(6)[2]	4:(6)[2]	7:(6)[2]	
M.6.22: [2]	3:(6)[2]			
M.6.23: [2]	8:(5)[2]			
M.6.24				
G.SP.M.6.B				
M.6.25: [2]	5:(6)[2]			
M.6.26				
M.6.27				
G.SP.M.6.C				
M.6.28: [1]	2:(6)[1]	6:(6)[2]		
M.6.29: [2]	5:(1)[2]			
MHM				
MHM1: [3]	11:(1)[1]	32:(6)[3]		
MHM2				
MHM3				
MHM4: [3]	3:(4)[2]			
MHM5: [3]	7:(1)[2]	8:(1)[2]		
MHM6: [2]	7:(1)[2]	13:(2)[2]		
MHM7				
MHM8				

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 6B2 MATH

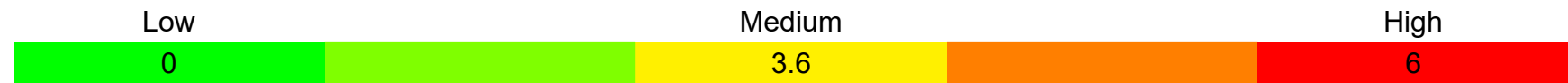
[illegible]

Table 6.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 6B2 MATH

Low		Medium		High	
1.2		3.6		6	
1	10199-1998	Exact:6			
2	10199-2306	Minimal:1		Negligible:5	
3	10199-1868	Exact:6			
4	REP10199-2013	Exact:6			
5	REP10199-2033	Exact:6			
6	REP10199-2012	Exact:6			
7	10199-2191	Exact:6			
8	10199-2145	Exact:6			
9	10199-1923	Exact:6			
10	10199-1785	Exact:6			
11	10199-1788	Partial:6			
12	10199-2078	Exact:6			
13	REP10199-10156	Exact:6			
14	10199-13345	Exact:6			
15	REP10199-10208	Exact:3		Partial:3	
16	10199-1980	Exact:6			
17	10199-1836	Partial:6			
18	10199-1887	Exact:6			
19	10199-2106	Exact:2		Partial:4	
20	10199-1546	Exact:6			
21	10199-1568	Exact:5		Partial:1	
22	10199-1872	Exact:5		Partial:1	
23	10199-2130	Exact:6			
24	10199-1463	Exact:6			
25	10199-1938	Exact:6			
26	10199-1962	Exact:6			
27	10199-1941	Exact:6			
28	REP10199-1854	Partial:2		Minimal:4	
29	10199-2125	Minimal:1		Negligible:5	
30	10199-1589	Exact:6			
31	REP10199-1956	Exact:6			
32	10199-2004	Exact:6			
33	10199-2118	Exact:6			
34	10199-1819	Exact:6			

Batch 3 West Virginia Math Grade 6

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B3_v2 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	1 2	6 5	54.55 45.45	15.17	0.41	YES
EE.M.6 Expressions and Equations	3	9	1 2	5 4	55.56 44.44	12	0.63	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	1 2	3 6	33.33 66.67	6.83	0.75	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0.17	0.41	NO
Total	18	37	1 2 3	14 17 6	38 46 16	34.17	0.41	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B3_v2 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	15.17	0.41	16.53	8	72.5	7	10.97	5	YES
EE.M.6 Expressions and Equations	3	9	12	0.63	30.4	9	58.3	17	11.3	15	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	6.83	0.75	26.49	13	71.13	17	2.38	6	YES
MHM Mathematical Habits of Mind	8	8	0.17	0.41	0	NaN	100	NaN	0	NaN	YES
Total	18	37	34.17	0.41	23.41	9	67.32	9.7	9.27	3.4	

Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B3_v2 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	15.17	0.41	9.83	0.75	89.39	6.84	YES	44	1	0.79	0.03	YES
EE.M.6 Expressions and Equations	3	9	12	0.63	6.5	0.55	72.22	6.09	YES	35	2	0.83	0.04	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	6.83	0.75	6.33	0.52	70.37	5.74	YES	20	2	0.94	0.06	YES
MHM Mathematical Habits of Mind	8	8	0.17	0.41	0.17	0.41	2.08	5.1	NO	0	1	N/A	0	NT
Total	18	37	34.17	0.41	5.7	4.03	58.52	39		25	19	0.85	0.1	

Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 6 B3_v2 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	YES	YES	YES	YES
EE.M.6 Expressions and Equations	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	YES	NO	NT

Table 6.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 6 B3_v2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	2	2	1	1	1
2	2	2	2	1	2	2
3	1	1	1	1	1	1
4	1	1	1	1	2	1
5	1	1	1	2	1	1
6	2	2	2	1	1	1
7	2	2	2	2	2	2
8	2	2	2	2	1	2
9	1	1	1	1	1	1
10	1	1	1	1	1	2
11	1	1	1	1	1	1
12	1	1	1	1	1	1
13	2	2	2	2	2	2
14	2	1	2	2	1	2
15	1	2	2	2	1	2
16	2	2	2	2	1	2
17	1	2	1	1	1	1
18	1	1	1	1	1	1
19	1	1	2	2	1	2
20	1	1	1	1	1	1
21	1	1	2	1	1	2
22	2	2	2	2	2	2
23	2	2	2	1	1	1
24	1	1	1	1	1	1
25	1	1	1	2	1	2
26	1	2	1	2	1	2
27	2	2	2	1	1	1
28	1	1	1	2	2	1
29	1	1	1	1	1	1
30	1	1	1	2	1	1
31	1	1	1	1	1	1
32	2	2	2	2	2	1
33	2	2	2	2	2	2
34	1	1	2	1	1	1

Intraclass correlation - .8495

Pairwise Comparison - 0.72

Table 6.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 6 B3_v2 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.6.24			2	M.6.24			2	M.6.24			1	M.6.24			1	M.6.24			1	M.6.24		
2	2	M.6.17			2	M.6.21			2	M.6.21			1	M.6.17			2	M.6.21			2	M.6.21		
3	1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28			1	M.6.28		
4	1	M.6.26			1	M.6.28			1	M.6.28			1	M.6.29			2	M.6.26			1	M.6.26		
5	1	M.6.22			1	M.6.22			1	M.6.22			2	M.6.22			1	M.6.22			1	M.6.22		
6	2	M.6.17			2	M.6.17			2	M.6.17			1	M.6.17			1	M.6.17			1	M.6.21		
7	2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25		
8	2	M.6.23			2	M.6.23			2	M.6.23			2	M.6.23			1	M.6.23			2	M.6.23		
9	1	M.6.18			1	M.6.18			1	M.6.18			1	M.6.18			1	M.6.18			1	M.6.18		
10	1	M.6.14			1	M.6.14			1	M.6.14			1	M.6.14			1	M.6.14			2	M.6.14		
11	1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1		
12	1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2		
13	2	M.6.20			2	M.6.18			2	M.6.19			2	M.6.20			2	M.6.18			2	M.6.18		
14	2	M.6.3			1	M.6.3			2	M.6.3			2	M.6.3			1	M.6.3			2	M.6.3		
15	1	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			1	M.6.18			2	M.6.16		
16	2	M.6.11			2	M.6.11			2	M.6.11			2	M.6.11			1	M.6.11			2	M.6.11		
17	1	M.6.4			2	M.6.4			1	M.6.3			1	M.6.4			1	M.6.4			1	M.6.4		
18	1	M.6.19			1	M.6.19			1	M.6.19			1	M.6.19			1	M.6.19			1	M.6.19		
19	1	M.6.15			1	M.6.14			2	M.6.14			2	M.6.15			1	M.6.14			2	M.6.15		
20	1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5			1	M.6.5		
21	1	M.6.17			1	M.6.17			2	M.6.17			1	M.6.17			1	M.6.17			2	M.6.17		
22	2	M.6.3			2	M.6.3	MHM6		2	M.6.3			2	M.6.3			2	M.6.3			2	M.6.3		
23	2	M.6.9			2	M.6.9			2	M.6.9			1	M.6.11			1	M.6.9			1	M.6.9		
24	1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6		
25	1	M.6.12			1	M.6.12			1	M.6.12			2	M.6.12			1	M.6.12			2	M.6.12		
26	1	M.6.1			2	M.6.1			1	M.6.1			2	M.6.3			1	M.6.1			2	M.6.1		
27	2	M.6.9			2	M.6.9			2	M.6.9			1	M.6.11			1	M.6.9			1	M.6.9		
28	1	M.6.7			1	M.6.7			1	M.6.7			2	M.6.7			2	M.6.7			1	M.6.7		
29	1	M.6.12			1	M.6.12			1	M.6.12			1	M.6.12			1	M.6.12			1	M.6.12		
30	1	M.6.15			1	M.6.15			1	M.6.15			2	M.6.15			1	M.6.15			1	M.6.15		
31	1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2		
32	2	M.6.8			2	M.6.18			2	M.6.18			2	M.6.18			2	M.6.18			1	M.6.18		
33	2	M.6.3			2	M.6.3			2	M.6.3			2	M.6.3			2	M.6.3			2	M.6.3		
34	1	M.6.10			1	M.6.10			2	M.6.10			1	M.6.10			1	M.6.10			1	M.6.10		
Objective Pairwise Comparison: 0.85																								
Standard Pairwise Comparison: 0.96																								

Table 6.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV MATH 2019 Grade 6 B3_v2

	Low	Medium		High	
	0	3.6		6	
RP.NS.M.6					
RP.NS.M.6A					
M.6.1	11(6)	26(5)			
M.6.2	12(6)	31(6)			
M.6.3	33(6)	14(6)	26(1)	17(1)	22(6)
RP.NS.M.6B					
M.6.4	17(5)				
RP.NS.M.6C					
M.6.5	20(6)				
M.6.6	24(6)				
M.6.7	28(6)				
RP.NS.M.6D					
M.6.8	32(1)				
M.6.9	27(5)	23(5)			
M.6.10	34(6)				
M.6.11	23(1)	16(6)	27(1)		
EE.M.6					
EE.M.6A					
M.6.12	25(6)	29(6)			
M.6.13					
M.6.14	19(3)	10(6)			
M.6.15	19(3)	30(6)			
EE.M.6B					

M.6.16	15(1)				
M.6.17	2(2)	6(5)	21(6)		
M.6.18	15(5)	9(6)	13(3)	32(5)	
M.6.19	13(1)	18(6)			
EE.M.6C					
M.6.20	13(2)				
G.SP.M.6					
G.SP.M.6.A					
M.6.21	6(1)	2(4)			
M.6.22	5(6)				
M.6.23	8(6)				
M.6.24	1(6)				
G.SP.M.6.B					
M.6.25	7(6)				
M.6.26	4(3)				
M.6.27					
G.SP.M.6.C					
M.6.28	4(2)	3(6)			
M.6.29	4(1)				
MHM					
MHM1					
MHM2					
MHM3					
MHM4					
MHM5					
MHM6	22(1)				
MHM7					
MHM8					

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

WV MATH 2019 Grade 6 B3_v2

Low		Medium		High	
1.2		3.6		6	
1	10199-1955	M.6.24:6			
2	REP10199-2013	M.6.17:2	M.6.21:4		
3	10199-2121	M.6.28:6			
4	10199-2417	M.6.26:3	M.6.28:2	M.6.29:1	
5	REP10199-2005	M.6.22:6			
6	REP10199-2008	M.6.17:5	M.6.21:1		
7	REP10199-2033	M.6.25:6			
8	10199-9847	M.6.23:6			
9	REP10199-11406	M.6.18:6			
10	10199-2057	M.6.14:6			
11	10199-9877	M.6.1:6			
12	10199-1961	M.6.2:6			
13	10199-1945	M.6.18:3	M.6.19:1	M.6.20:2	
14	10199-10121	M.6.3:6			
15	REP10199-1854	M.6.16:1	M.6.18:5		
16	10199-1765	M.6.11:6			
17	REP10199-10208	M.6.3:1	M.6.4:5		
18	10199-1791	M.6.19:6			
19	10199-10153	M.6.14:3	M.6.15:3		
20	10199-1538	M.6.5:6			
21	10199-10194	M.6.17:6			
22	REP10199-10156	M.6.3:6	MHM6:1		
23	10199-10851	M.6.9:5	M.6.11:1		
24	10199-1584	M.6.6:6			
25	10199-1975	M.6.12:6			
26	10199-2073	M.6.1:5	M.6.3:1		
27	10199-2090	M.6.9:5	M.6.11:1		
28	10199-1953	M.6.7:6			
29	10199-1939	M.6.12:6			
30	10199-1933	M.6.15:6			
31	10199-1960	M.6.2:6			
32	REP10199-1887	M.6.8:1	M.6.18:5		
33	10199-1767	M.6.3:6			
34	10199-1708	M.6.10:6			

Table 6.9

Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])

WV MATH 2019 Grade 6 B3_v2

Low DOK		Matched DOK		High DOK

RP.NS.M.6					
RP.NS.M.6A					
M.6.1: [1]	11:(6)[1]	26:(5)[1]			
M.6.2: [1]	12:(6)[1]	31:(6)[1]			
M.6.3: [2]	14:(6)[2]	17:(1)[1]	22:(6)[2]	26:(1)[2]	33:(6)[2]
RP.NS.M.6B					
M.6.4: [2]	17:(5)[1]				
RP.NS.M.6C					
M.6.5: [1]	20:(6)[1]				
M.6.6: [1]	24:(6)[1]				
M.6.7: [1]	28:(6)[1]				
RP.NS.M.6D					
M.6.8: [2]	32:(1)[2]				
M.6.9: [1]	23:(5)[2]	27:(5)[2]			
M.6.10: [2]	34:(6)[1]				
M.6.11: [2]	16:(6)[2]	23:(1)[1]	27:(1)[1]		
EE.M.6					
EE.M.6A					
M.6.12: [1]	25:(6)[1]	29:(6)[1]			
M.6.13					
M.6.14: [1]	10:(6)[1]	19:(3)[1]			
M.6.15: [1]	19:(3)[2]	30:(6)[1]			
EE.M.6B					
M.6.16: [1]	15:(1)[2]				
M.6.17: [2]	2:(2)[2]	6:(5)[2]	21:(6)[1]		
M.6.18: [2]	9:(6)[1]	13:(3)[2]	15:(5)[2]	32:(5)[2]	
M.6.19: [2]	13:(1)[2]	18:(6)[1]			
EE.M.6C					

M.6.20: [2]	13:(2)[2]				
G.SP.M.6					
G.SP.M.6.A					
M.6.21: [2]	2:(4)[2]	6:(1)[1]			
M.6.22: [2]	5:(6)[1]				
M.6.23: [2]	8:(6)[2]				
M.6.24: [2]	1:(6)[2]				
G.SP.M.6.B					
M.6.25: [2]	7:(6)[2]				
M.6.26: [1]	4:(3)[1]				
M.6.27					
G.SP.M.6.C					
M.6.28: [1]	3:(6)[1]	4:(2)[1]			
M.6.29: [2]	4:(1)[1]				
MHM					
MHM1					
MHM2					
MHM3					
MHM4					
MHM5					
MHM6: [2]	22:(1)[2]				
MHM7					
MHM8					

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 6B3 MATH

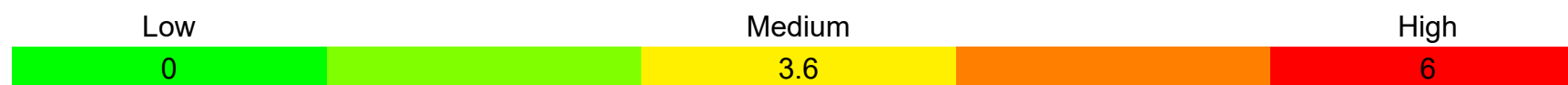
[illegible]

Table 6.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 6B3 MATH

Low		Medium		High	
1.2		3.6		6	
1	10199-1955	Exact:5		Partial:1	
2	REP10199-2013	Exact:6			
3	10199-2121	Exact:5		Partial:1	
4	10199-2417	Exact:2		Partial:4	
5	REP10199-2005	Exact:5		Partial:1	
6	REP10199-2008	Partial:5		Minimal:1	
7	REP10199-2033	Exact:6			
8	10199-9847	Exact:6			
9	REP10199-11406	Exact:6			
10	10199-2057	Exact:6			
11	10199-9877	Exact:6			
12	10199-1961	Exact:6			
13	10199-1945	Exact:3		Partial:3	
14	10199-10121	Exact:6			
15	REP10199-1854	Partial:1		Minimal:5	
16	10199-1765	Exact:6			
17	REP10199-10208	Exact:3		Partial:3	
18	10199-1791	Exact:6			
19	10199-10153	Exact:4		Partial:2	
20	10199-1538	Exact:6			
21	10199-10194	Exact:6			
22	REP10199-10156	Exact:6			
23	10199-10851	Exact:5		Minimal:1	
24	10199-1584	Exact:6			
25	10199-1975	Exact:6			
26	10199-2073	Exact:5		Partial:1	
27	10199-2090	Exact:5		Partial:1	
28	10199-1953	Exact:6			
29	10199-1939	Exact:6			
30	10199-1933	Exact:6			
31	10199-1960	Exact:6			
32	REP10199-1887	Exact:5		Partial:1	
33	10199-1767	Exact:6			
34	10199-1708	Exact:6			

Batch 4 West Virginia Math Grade 6

Table 6.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B4_v2 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	1 2	6 5	54.55 45.45	15.33	1.03	YES
EE.M.6 Expressions and Equations	3	9	1 2	5 4	55.56 44.44	13.33	0.82	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	1 2	3 6	33.33 66.67	6.83	0.41	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	1.17	0.41	NO
Total	18	37	1 2 3	14 17 6	38 46 16	36.66	1.21	

Table 6.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B4_v2 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	15.33	1.03	18.83	13	62.8	7	18.38	10	YES
EE.M.6 Expressions and Equations	3	9	13.33	0.82	24.62	9	62.74	8	12.65	6	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	6.83	0.41	29.76	23	62.7	25	7.54	12	YES
MHM Mathematical Habits of Mind	8	8	1.17	0.41	0	0	100	0	0	0	YES
Total	18	37	36.66	1.21	22.27	7	64.09	4.8	13.64	6	

Table 6.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 6 B4_v2 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	4	11	15.33	1.03	9.5	0.84	86.36	7.61	YES	40	2	0.8	0.02	YES
EE.M.6 Expressions and Equations	3	9	13.33	0.82	8.33	0.82	92.59	9.07	YES	38	2	0.8	0.04	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	3	9	6.83	0.41	6.67	0.52	74.07	5.74	YES	19	1	0.98	0.05	YES
MHM Mathematical Habits of Mind	8	8	1.17	0.41	1.17	0.41	14.58	5.1	NO	3	1	0.92	0.2	YES
Total	18	37	36.66	1.21	6.4	3.69	66.9	36		25	17	0.88	0.09	

Table 6.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 6 B4_v2 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
RP.NS.M.6 Ratios and Proportional Relationships and the Number System	YES	YES	YES	YES
EE.M.6 Expressions and Equations	YES	YES	YES	YES
G.SP.M.6 Geometry (G) and Statistics and Probability	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	YES	NO	YES

Table 6.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 6 B4_v2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	2	2	2	1	1
2	2	2	2	1	2	1
3	2	1	2	1	1	1
4	2	1	1	2	1	1
5	1	1	1	1	2	1
6	2	2	2	2	2	2
7	2	1	2	2	2	1
8	2	1	2	2	1	2
9	1	1	1	1	1	1
10	2	1	2	1	1	1
11	1	1	1	1	1	1
12	2	2	2	2	2	2
13	2	2	2	2	2	2
14	2	1	1	1	2	2
15	2	2	2	1	2	2
16	2	2	2	1	1	2
17	1	1	2	1	1	2
18	1	1	1	1	1	2
19	2	2	2	1	2	2
20	1	1	2	1	1	1
21	1	1	1	1	1	1
22	1	1	1	1	1	2
23	1	1	1	1	2	2
24	2	2	2	2	2	2
25	1	1	1	1	1	1
26	1	1	1	1	1	1
27	1	1	1	1	1	1
28	1	1	1	1	1	2
29	2	2	2	1	1	2
30	1	2	1	1	1	1
31	1	1	1	1	1	1
32	1	1	1	1	1	1
33	3	3	3	3	3	3
34	1	1	1	1	1	1

Intraclass correlation - .9101

Pairwise Comparison - 0.73

Table 6.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 6 B4_v2 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.6.24			2	M.6.24			2	M.6.24			2	M.6.24			1	M.6.24			1	M.6.24		
2	2	M.6.17			2	M.6.17			2	M.6.17			1	M.6.17			2	M.6.17			1	M.6.21		
3	2	M.6.26			1	M.6.26			2	M.6.29			1	M.6.26			1	M.6.26			1	M.6.26		
4	2	M.6.26			1	M.6.22			1	M.6.22			2	M.6.22			1	M.6.22			1	M.6.22		
5	1	M.6.22			1	M.6.28			1	M.6.28			1	M.6.28			2	M.6.28			1	M.6.28		
6	2	M.6.21			2	M.6.21			2	M.6.21			2	M.6.17	M.6.21		2	M.6.17			2	M.6.13		
7	2	M.6.25			1	M.6.25			2	M.6.25			2	M.6.25			2	M.6.25			1	M.6.25		
8	2	M.6.23			1	M.6.23			2	M.6.23			2	M.6.23			1	M.6.23			2	M.6.23		
9	1	M.6.14			1	M.6.14			1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.14		
10	2	M.6.19			1	M.6.19			2	M.6.19			1	M.6.19			1	M.6.19			1	M.6.19		
11	1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.11		
12	2	M.6.3			2	M.6.3	MHM6		2	M.6.3			2	M.6.3			2	M.6.3			2	M.6.3		
13	2	M.6.14			2	M.6.14			2	M.6.14			2	M.6.14			2	M.6.14			2	M.6.14		
14	2	M.6.3			1	M.6.1			1	M.6.1			1	M.6.3			2	M.6.3			2	M.6.3		
15	2	M.6.20			2	M.6.20			2	M.6.20			1	M.6.20			2	M.6.20			2	M.6.20		
16	2	M.6.11			2	M.6.11			2	M.6.11			1	M.6.11			1	M.6.11			2	M.6.6		
17	1	M.6.18			1	M.6.18			2	M.6.18			1	M.6.16			1	M.6.16			2	M.6.16		
18	1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			1	M.6.6			2	M.6.6		
19	2	M.6.4			2	M.6.4			2	M.6.4			1	M.6.4			2	M.6.4			2	M.6.4		
20	1	M.6.14			1	M.6.14			2	M.6.14			1	M.6.14			1	M.6.4			1	M.6.14		
21	1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1			1	M.6.1		
22	1	M.6.18			1	M.6.18			1	M.6.18			1	M.6.18			1	M.6.18			2	M.6.18		
23	1	M.6.10			1	M.6.10			1	M.6.10			1	M.6.19			2	M.6.10			2	M.6.10		
24	2	M.6.17			2	M.6.13			2	M.6.17			2	M.6.17			2	M.6.17			2	M.6.17		
25	1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7			1	M.6.7		
26	1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2			1	M.6.2		
27	1	M.6.12			1	M.6.12			1	M.6.13			1	M.6.12			1	M.6.12			1	M.6.12		
28	1	M.6.8			1	M.6.8			1	M.6.8			1	M.6.8			1	M.6.8			2	M.6.10		
29	2	M.6.9			2	M.6.9	M.6.11		2	M.6.9			1	M.6.11			1	M.6.9			2	M.6.9		
30	1	M.6.1			2	M.6.1			1	M.6.1			1	M.6.1			1	M.6.3			1	M.6.1		
31	1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15			1	M.6.15		
32	1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13			1	M.6.13		
33	3	M.6.3	MHM1		3	M.6.3	MHM1		3	M.6.3	MHM1		3	M.6.3	MHM1		3	M.6.3	MHM1		3	M.6.3	MHM1	
34	1	M.6.18			1	M.6.18			1	M.6.18			1	M.6.18			1	M.6.18			1	M.6.18		
Objective Pairwise Comparison: 0.78																								
Standard Pairwise Comparison: 0.93																								

Table 6.7

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)

WV MATH 2019 Grade 6 B4_v2

Low		Medium		High
0		6		10
RP.NS.M.6				
RP.NS.M.6A				
M.6.1	14(2)	21(6)	30(5)	
M.6.2	26(6)	11(5)		
M.6.3	30(1)	14(4)	12(6)	33(6)
RP.NS.M.6B				
M.6.4	19(6)	20(1)		
RP.NS.M.6C				
M.6.5				
M.6.6	18(6)	16(1)		
M.6.7	25(6)			
RP.NS.M.6D				
M.6.8	28(5)			
M.6.9	29(10)			
M.6.10	28(1)	23(5)		
M.6.11	11(1)	16(5)	29(4)	
EE.M.6				
EE.M.6A				
M.6.12	27(5)			
M.6.13	27(1)	32(6)	24(1)	6(1)
M.6.14	13(6)	9(3)	20(5)	
M.6.15	31(6)	9(3)		
EE.M.6B				

M.6.16	17(3)			
M.6.17	6(2)	2(5)	24(5)	
M.6.18	22(6)	17(3)	34(6)	
M.6.19	10(6)	23(1)		
EE.M.6C				
M.6.20	15(6)			
G.SP.M.6				
G.SP.M.6.A				
M.6.21	2(1)	6(4)		
M.6.22	5(1)	4(5)		
M.6.23	8(6)			
M.6.24	1(6)			
G.SP.M.6.B				
M.6.25	7(6)			
M.6.26	4(1)	3(5)		
M.6.27				
G.SP.M.6.C				
M.6.28	5(5)			
M.6.29	3(1)			
MHM				
MHM1	33(6)			
MHM2				
MHM3				
MHM4				
MHM5				
MHM6	12(1)			
MHM7				
MHM8				

Table 6.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

WV MATH 2019 Grade 6 B4_v2

Low		Medium		High	
2		6		10	
1	10199-1789	M.6.24:6			
2	REP10199-2008	M.6.17:5	M.6.21:1		
3	10199-2304	M.6.26:5	M.6.29:1		
4	REP10199-2005	M.6.22:5	M.6.26:1		
5	10199-1795	M.6.22:1	M.6.28:5		
6	REP10199-2013	M.6.13:1	M.6.17:2	M.6.21:4	
7	10199-2034	M.6.25:6			
8	10199-2380	M.6.23:6			
9	10199-12573	M.6.14:3	M.6.15:3		
10	10199-2064	M.6.19:6			
11	10199-1808	M.6.2:5	M.6.11:1		
12	REP10199-10156	M.6.3:6	MHM6:1		
13	10199-12638	M.6.14:6			
14	10199-1948	M.6.1:2	M.6.3:4		
15	10199-1886	M.6.20:6			
16	10199-1766	M.6.6:1	M.6.11:5		
17	10199-1858	M.6.16:3	M.6.18:3		
18	10199-1587	M.6.6:6			
19	10199-1687	M.6.4:6			
20	10199-2055	M.6.4:1	M.6.14:5		
21	10199-9843	M.6.1:6			
22	10199-2037	M.6.18:6			
23	10199-1709	M.6.10:5	M.6.19:1		
24	10199-1864	M.6.13:1	M.6.17:5		
25	10199-1560	M.6.7:6			
26	10199-1811	M.6.2:6			
27	10199-2052	M.6.12:5	M.6.13:1		
28	10199-1702	M.6.8:5	M.6.10:1		
29	10199-2092	M.6.9:10	M.6.11:4		
30	10199-9802	M.6.1:5	M.6.3:1		
31	10199-2426	M.6.15:6			
32	10199-1915	M.6.13:6			
33	10199-1976	M.6.3:6	MHM1:6		
34	REP10199-11406	M.6.18:6			

Table 6.9

Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])

WV MATH 2019 Grade 6 B4_v2

Low DOK		Matched DOK		High DOK

RP.NS.M.6				
RP.NS.M.6A				
M.6.1: [1]	14:(2)[1]	21:(6)[1]	30:(5)[1]	
M.6.2: [1]	11:(5)[1]	26:(6)[1]		
M.6.3: [2]	12:(6)[2]	14:(4)[2]	30:(1)[1]	33:(6)[3]
RP.NS.M.6B				
M.6.4: [2]	19:(6)[2]	20:(1)[1]		
RP.NS.M.6C				
M.6.5				
M.6.6: [1]	16:(1)[2]	18:(6)[1]		
M.6.7: [1]	25:(6)[1]			
RP.NS.M.6D				
M.6.8: [2]	28:(5)[1]			
M.6.9: [1]	29:(10)[2]			
M.6.10: [2]	23:(5)[1]	28:(1)[2]		
M.6.11: [2]	11:(1)[1]	16:(5)[2]	29:(4)[2]	
EE.M.6				
EE.M.6A				
M.6.12: [1]	27:(5)[1]			
M.6.13: [1]	6:(1)[2]	24:(1)[2]	27:(1)[1]	32:(6)[1]
M.6.14: [1]	9:(3)[1]	13:(6)[2]	20:(5)[1]	
M.6.15: [1]	9:(3)[1]	31:(6)[1]		
EE.M.6B				
M.6.16: [1]	17:(3)[1]			
M.6.17: [2]	2:(5)[2]	6:(2)[2]	24:(5)[2]	
M.6.18: [2]	17:(3)[1]	22:(6)[1]	34:(6)[1]	
M.6.19: [2]	10:(6)[1]	23:(1)[1]		
EE.M.6C				

M.6.20: [2]	15:(6)[2]			
G.SP.M.6				
G.SP.M.6.A				
M.6.21: [2]	2:(1)[1]	6:(4)[2]		
M.6.22: [2]	4:(5)[1]	5:(1)[1]		
M.6.23: [2]	8:(6)[2]			
M.6.24: [2]	1:(6)[2]			
G.SP.M.6.B				
M.6.25: [2]	7:(6)[2]			
M.6.26: [1]	3:(5)[1]	4:(1)[2]		
M.6.27				
G.SP.M.6.C				
M.6.28: [1]	5:(5)[1]			
M.6.29: [2]	3:(1)[2]			
MHM				
MHM1: [3]	33:(6)[3]			
MHM2				
MHM3				
MHM4				
MHM5				
MHM6: [2]	12:(1)[2]			
MHM7				
MHM8				

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 6B4 MATH

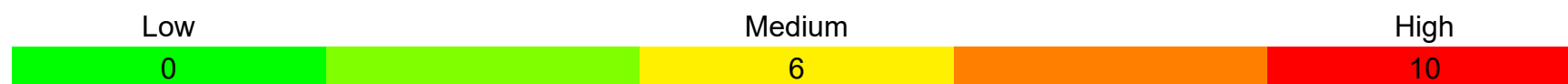
[illegible]

Table 6.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 6B4 MATH

Low		Medium		High	
2		6		10	
1	10199-1789	Exact:6			
2	REP10199-2008	Partial:5	Minimal:1		
3	10199-2304	Exact:5	Partial:1		
4	REP10199-2005	Exact:5	Partial:1		
5	10199-1795	Exact:6			
6	REP10199-2013	Exact:4	Partial:1	Minimal:1	
7	10199-2034	Exact:6			
8	10199-2380	Exact:6			
9	10199-12573	Exact:3	Partial:3		
10	10199-2064	Exact:6			
11	10199-1808	Exact:5	Partial:1		
12	REP10199-10156	Exact:6			
13	10199-12638	Exact:6			
14	10199-1948	Exact:4	Partial:2		
15	10199-1886	Exact:5	Partial:1		
16	10199-1766	Exact:6			
17	10199-1858	Exact:3	Minimal:3		
18	10199-1587	Exact:6			
19	10199-1687	Exact:6			
20	10199-2055	Exact:6			
21	10199-9843	Exact:6			
22	10199-2037	Exact:6			
23	10199-1709	Exact:5	Partial:1		
24	10199-1864	Exact:5	Partial:1		
25	10199-1560	Exact:6			
26	10199-1811	Exact:6			
27	10199-2052	Exact:5	Partial:1		
28	10199-1702	Exact:5	Partial:1		
29	10199-2092	Exact:10	Partial:2		
30	10199-9802	Exact:6			
31	10199-2426	Exact:6			
32	10199-1915	Exact:6			
33	10199-1976	Exact:6			
34	REP10199-11406	Exact:6			

Batch 1 West Virginia Math Grade 7

Table 7.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 7 B1 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	2	6	2	6	100	10.17	0.41	YES
EE.M.7 Expressions and Equations	2	4	1 2	1 3	25 75	8	0.63	YES
G.M.7 Geometry	2	6	2	6	100	7.83	0.41	YES
SP.M.7 Statistics and Probability	3	10	1 2 3	1 6 3	10 60 30	9	0	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	17	34	1 2 3	2 23 9	6 68 26	35	0	

Table 7.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 7 B1 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	2	6	10.17	0.41	60.91	16	39.09	16	0	0	NO
EE.M.7 Expressions and Equations	2	4	8	0.63	20.6	19	74.64	16	4.76	12	YES
G.M.7 Geometry	2	6	7.83	0.41	51.19	29	48.81	29	0	0	WEAK
SP.M.7 Statistics and Probability	3	10	9	0	40.74	9	55.56	10	3.7	6	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	17	34	35	0	44.29	15.1	53.81	14.2	1.9	3.5	

Table 7.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 7 B1 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	2	6	10.17	0.41	5	0.63	83.33	10.54	YES	27	1	0.8	0.01	YES
EE.M.7 Expressions and Equations	2	4	8	0.63	4	0	100	0	YES	24	2	0.85	0.04	YES
G.M.7 Geometry	2	6	7.83	0.41	6	0	100	0	YES	23	1	0.84	0.02	YES
SP.M.7 Statistics and Probability	3	10	9	0	5.5	0.55	55	5.48	YES	26	0	0.87	0.04	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	17	34	35	0	4.1	2.41	67.67	42		20	11	0.84	0.08	

Table 7.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 7 B1 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	YES	NO	YES	YES
EE.M.7 Expressions and Equations	YES	YES	YES	YES
G.M.7 Geometry	YES	WEAK	YES	YES
SP.M.7 Statistics and Probability	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 7.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 7 B1

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	2	2	2	1	2
2	2	1	2	1	1	1
3	2	1	2	2	1	2
4	2	1	2	1	1	1
5	2	2	2	2	1	1
6	1	1	1	1	1	1
7	2	2	2	2	1	2
8	2	2	2	2	1	2
9	1	1	2	2	2	1
10	2	1	2	1	2	2
11	1	1	2	1	1	1
12	2	1	2	2	1	2
13	2	1	2	2	1	2
14	2	1	2	2	1	2
15	1	1	2	1	1	1
16	2	2	2	2	2	2
17	1	1	1	1	1	1
18	2	2	2	2	2	2
19	2	1	2	2	1	2
20	2	2	2	2	2	1
21	1	1	2	1	1	1
22	2	2	2	2	2	2
23	2	1	2	2	2	1
24	2	2	2	2	2	2
25	1	1	1	1	1	1
26	1	1	2	1	1	1
27	2	2	2	2	1	2
28	1	1	1	1	1	1
29	2	2	2	2	2	2
30	1	1	1	1	1	1
31	1	1	1	1	1	1
32	1	1	1	1	1	1
33	1	1	1	1	1	1
34	2	2	2	2	2	2

Intraclass correlation - .9001

Pairwise Comparison - 0.74

Table 7.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 7 B1 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.7.6			2	M.7.6			2	M.7.4			2	M.7.6			1	M.7.6			2	M.7.6		
2	2	M.7.12			1	M.7.12			2	M.7.12			1	M.7.12			1	M.7.12			1	M.7.12		
3	2	M.7.9			1	M.7.9			2	M.7.6			2	M.7.9			1	M.7.9			2	M.7.6		
4	2	M.7.15			1	M.7.15			2	M.7.15			1	M.7.15			1	M.7.15			1	M.7.15		
5	2	M.7.24			2	M.7.24			2	M.7.24			2	M.7.24			1	M.7.24			1	M.7.24		
6	1	M.7.1			1	M.7.1			1	M.7.2			1	M.7.1			1	M.7.1			1	M.7.1		
7	2	M.7.8			2	M.7.8			2	M.7.8			2	M.7.8			1	M.7.8			2	M.7.8		
8	2	M.7.16			2	M.7.16			2	M.7.16			2	M.7.16			1	M.7.16			2	M.7.16		
9	1	M.7.23			1	M.7.23			2	M.7.23			2	M.7.26			2	M.7.23			1	M.7.23		
10	2	M.7.3			1	M.7.3			2	M.7.3			1	M.7.3			2	M.7.3			2	M.7.9		
11	1	M.7.7			1	M.7.7			2	M.7.7			1	M.7.7			1	M.7.7			1	M.7.7		
12	2	M.7.11			1	M.7.11			2	M.7.11			2	M.7.11			1	M.7.11			2	M.7.11		
13	2	M.7.4			1	M.7.4			2	M.7.4			2	M.7.4			1	M.7.4			2	M.7.4		
14	2	M.7.18			1	M.7.18			2	M.7.18			2	M.7.18			1	M.7.18			2	M.7.18		
15	1	M.7.10			1	M.7.10			2	M.7.10			1	M.7.10			1	M.7.10			1	M.7.10		
16	2	M.7.22			2	M.7.21			2	M.7.21			2	M.7.22			2	M.7.22			2	M.7.22		
17	1	M.7.15			1	M.7.15			1	M.7.15			1	M.7.15			1	M.7.15			1	M.7.15		
18	2	M.7.6			2	M.7.3			2	M.7.6			2	M.7.6			2	M.7.6			2	M.7.6		
19	2	M.7.12			1	M.7.12			2	M.7.12			2	M.7.8			1	M.7.12			2	M.7.12		
20	2	M.7.24			2	M.7.24			2	M.7.24			2	M.7.24			2	M.7.24			1	M.7.24		
21	1	M.7.7			1	M.7.8			2	M.7.7			1	M.7.7			1	M.7.7			1	M.7.7		
22	2	M.7.9			2	M.7.9			2	M.7.9			2	M.7.9			2	M.7.9			2	M.7.9		
23	2	M.7.18			1	M.7.18			2	M.7.18			2	M.7.18			2	M.7.18			1	M.7.18		
24	2	M.7.22			2	M.7.22			2	M.7.22			2	M.7.22			2	M.7.22			2	M.7.22		
25	1	M.7.1			1	M.7.1			1	M.7.1			1	M.7.1			1	M.7.1			1	M.7.1		
26	1	M.7.14			1	M.7.14			2	M.7.14			1	M.7.14			1	M.7.14			1	M.7.14		
27	2	M.7.8			2	M.7.8			2	M.7.8			2	M.7.8			1	M.7.8			2	M.7.8		
28	1	M.7.1			1	M.7.1			1	M.7.2			1	M.7.1			1	M.7.1			1	M.7.1		
29	2	M.7.13			2	M.7.13			2	M.7.13			2	M.7.13			2	M.7.13			2	M.7.13		
30	1	M.7.23			1	M.7.23			1	M.7.23			1	M.7.23			1	M.7.23			1	M.7.23		
31	1	M.7.7			1	M.7.7			1	M.7.7			1	M.7.7			1	M.7.7			1	M.7.7		
32	1	M.7.1			1	M.7.1			1	M.7.2			1	M.7.1			1	M.7.1			1	M.7.1		
33	1	M.7.5			1	M.7.5			1	M.7.5			1	M.7.5			1	M.7.5			1	M.7.5		
34	2	M.7.25			2	M.7.25			2	M.7.25			2	M.7.25			2	M.7.25			2	M.7.25		
Objective Pairwise Comparison: 0.88																								
Standard Pairwise Comparison: 0.96																								

Table 7.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV MATH 2019 Grade 7 B1

	Low	Medium		High
	0	7.2		12
RP.NS.M.7				
RP.NS.M.7A				
M.7.1	6(5)	25(6)	28(5)	32(5)
M.7.2	32(1)	28(1)	6(1)	
M.7.3	10(5)	18(1)		
RP.NS.M.7B				
M.7.4	1(1)	13(12)		
M.7.5	33(6)			
M.7.6	1(5)	3(2)	18(5)	
EE.M.7				
EE.M.7A				
M.7.7	21(5)	11(6)	31(6)	
M.7.8	27(6)	7(6)	21(1)	19(1)
EE.M.7B				
M.7.9	22(6)	10(1)	3(4)	
M.7.10	15(6)			
G.M.7				
G.M.7.A				
M.7.11	12(6)			
M.7.12	2(6)	19(5)		
M.7.13	29(6)			
G.M.7.B				

M.7.14	26(6)			
M.7.15	17(6)	4(6)		
M.7.16	8(6)			
SP.M.7				
SP.M.7.A				
M.7.17				
M.7.18	14(6)	23(6)		
SP.M.7.B				
M.7.19				
M.7.20				
M.7.21	16(2)			
M.7.22	16(4)	24(6)		
SP.M.7.C				
M.7.23	30(6)	9(5)		
M.7.24	5(6)	20(6)		
M.7.25	34(6)			
M.7.26	9(1)			
MHM				
MHM1				
MHM2				
MHM3				
MHM4				
MHM5				
MHM6				
MHM7				
MHM8				

Table 7.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV MATH 2019 Grade 7 B1

Low		Medium		High	
2.4		7.2		12	
1	10199-2273	M.7.4:1		M.7.6:5	
2	10199-2243	M.7.12:6			
3	10199-2183	M.7.6:2		M.7.9:4	
4	10199-2155	M.7.15:6			
5	REP10199-2448	M.7.24:6			
6	10199-2229	M.7.1:5		M.7.2:1	
7	10199-2509	M.7.8:6			
8	10199-2200	M.7.16:6			
9	REP10199-2585	M.7.23:5		M.7.26:1	
10	10199-2163	M.7.3:5		M.7.9:1	
11	10199-2285	M.7.7:6			
12	10199-2441	M.7.11:6			
13	REP10199-2530	M.7.4:12			
14	10199-2607	M.7.18:6			
15	10199-2138	M.7.10:6			
16	10199-2567	M.7.21:2		M.7.22:4	
17	10199-2254	M.7.15:6			
18	10199-2325	M.7.3:1		M.7.6:5	
19	REP10199-2662	M.7.8:1		M.7.12:5	
20	10199-2460	M.7.24:6			
21	REP10199-2428	M.7.7:5		M.7.8:1	
22	REP10199-2299	M.7.9:6			
23	10199-2609	M.7.18:6			
24	REP10199-2616	M.7.22:6			
25	10199-2601	M.7.1:6			
26	10199-2245	M.7.14:6			
27	REP10199-2175	M.7.8:6			
28	10199-2208	M.7.1:5		M.7.2:1	
29	REP10199-2950	M.7.13:6			
30	10199-2445	M.7.23:6			
31	10199-10452	M.7.7:6			
32	10199-2543	M.7.1:5		M.7.2:1	
33	10199-2140	M.7.5:6			
34	10199-2619	M.7.25:6			

Table 7.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
 WV MATH 2019 Grade 7 B1

Low DOK		Matched DOK		High DOK

RP.NS.M.7				
RP.NS.M.7A				
M.7.1: [2]	6:(5)[1]	25:(6)[1]	28:(5)[1]	32:(5)[1]
M.7.2: [2]	6:(1)[1]	28:(1)[1]	32:(1)[1]	
M.7.3: [2]	10:(5)[2]	18:(1)[2]		
RP.NS.M.7B				
M.7.4: [2]	1:(1)[2]	13:(12)[2]		
M.7.5: [2]	33:(6)[1]			
M.7.6: [2]	1:(5)[2]	3:(2)[2]	18:(5)[2]	
EE.M.7				
EE.M.7A				
M.7.7: [1]	11:(6)[1]	21:(5)[1]	31:(6)[1]	
M.7.8: [2]	7:(6)[2]	19:(1)[2]	21:(1)[1]	27:(6)[2]
EE.M.7B				
M.7.9: [2]	3:(4)[2]	10:(1)[2]	22:(6)[2]	
M.7.10: [2]	15:(6)[1]			
G.M.7				
G.M.7.A				
M.7.11: [2]	12:(6)[2]			
M.7.12: [2]	2:(6)[1]	19:(5)[2]		
M.7.13: [2]	29:(6)[2]			
G.M.7.B				
M.7.14: [2]	26:(6)[1]			
M.7.15: [2]	4:(6)[1]	17:(6)[1]		
M.7.16: [2]	8:(6)[2]			
SP.M.7				
SP.M.7.A				
M.7.17				

M.7.18: [3]	14:(6)[2]	23:(6)[2]		
SP.M.7.B				
M.7.19				
M.7.20				
M.7.21: [2]	16:(2)[2]			
M.7.22: [2]	16:(4)[2]	24:(6)[2]		
SP.M.7.C				
M.7.23: [1]	9:(5)[1]	30:(6)[1]		
M.7.24: [2]	5:(6)[2]	20:(6)[2]		
M.7.25: [3]	34:(6)[2]			
M.7.26: [3]	9:(1)[2]			
MHM				
MHM1				
MHM2				
MHM3				
MHM4				
MHM5				
MHM6				
MHM7				
MHM8				

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 7B1 MATH

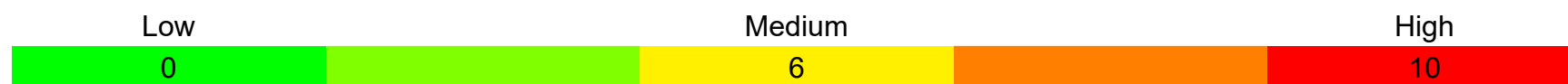
[illegible]

Table 7.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 7B1 MATH

Low	Medium	High
2	6	10

1 10199-2273	Exact:5	Partial:1
2 10199-2243	Exact:5	Partial:1
3 10199-2183	Exact:3	Partial:3
4 10199-2155	Exact:6	
5 REP10199-2448	Exact:6	
6 10199-2229	Exact:5	Partial:1
7 10199-2509	Exact:6	
8 10199-2200	Exact:6	
9 REP10199-2585	Exact:5	Partial:1
10 10199-2163	Exact:6	
11 10199-2285	Exact:6	
12 10199-2441	Exact:6	
13 REP10199-2530	Exact:10	Partial:2
14 10199-2607	Exact:5	Partial:1
15 10199-2138	Exact:5	Partial:1
16 10199-2567	Exact:4	Partial:2
17 10199-2254	Exact:6	
18 10199-2325	Exact:5	Partial:1
19 REP10199-2662	Exact:5	Partial:1
20 10199-2460	Exact:5	Partial:1
21 REP10199-2428	Exact:4	Partial:2
22 REP10199-2299	Exact:5	Partial:1
23 10199-2609	Exact:5	Partial:1
24 REP10199-2616	Exact:6	
25 10199-2601	Exact:6	
26 10199-2245	Exact:5	Partial:1
27 REP10199-2175	Exact:6	
28 10199-2208	Exact:5	Partial:1
29 REP10199-2950	Exact:6	
30 10199-2445	Exact:5	Partial:1
31 10199-10452	Exact:6	
32 10199-2543	Exact:5	Partial:1
33 10199-2140	Exact:6	
34 10199-2619	Exact:5	Partial:1

Batch 2 West Virginia Math Grade 7

Table 7.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 7 B2 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	2	6	2	6	100	11.33	0.82	YES
EE.M.7 Expressions and Equations	2	4	1 2	1 3	25 75	9.67	0.82	YES
G.M.7 Geometry	2	6	2	6	100	9	0	YES
SP.M.7 Statistics and Probability	3	10	1 2 3	1 6 3	10 60 30	8	0	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0.17	0.41	NO
Total	17	34	1 2 3	2 23 9	6 68 26	38.17	0.41	

Table 7.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 7 B2 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	2	6	11.33	0.82	30.42	18	69.58	18	0	0	YES
EE.M.7 Expressions and Equations	2	4	9.67	0.82	15	21	74.58	22	10.42	11	YES
G.M.7 Geometry	2	6	9	0	22.22	14	77.78	14	0	0	YES
SP.M.7 Statistics and Probability	3	10	8	0	31.25	10	58.33	10	10.42	5	YES
MHM Mathematical Habits of Mind	8	8	0.17	0.41	0	NaN	100	NaN	0	NaN	YES
Total	17	34	38.17	0.41	24.45	13.5	70.74	13.6	4.8	3.1	

Table 7.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 7 B2 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	2	6	11.33	0.82	4.83	0.41	80.56	6.8	YES	27	2	0.81	0.05	YES
EE.M.7 Expressions and Equations	2	4	9.67	0.82	4	0	100	0	YES	25	2	0.9	0.01	YES
G.M.7 Geometry	2	6	9	0	6	0	100	0	YES	23	0	0.78	0	YES
SP.M.7 Statistics and Probability	3	10	8	0	6	0	60	0	YES	23	0	0.83	0	YES
MHM Mathematical Habits of Mind	8	8	0.17	0.41	0.17	0.41	2.08	5.1	NO	0	1	N/A	0	NT
Total	17	34	38.17	0.41	4.2	2.41	68.53	41		20	11	0.83	0.09	

Table 7.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 7 B2 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	YES	YES	YES	YES
EE.M.7 Expressions and Equations	YES	YES	YES	YES
G.M.7 Geometry	YES	YES	YES	YES
SP.M.7 Statistics and Probability	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	YES	NO	NT

Table 7.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 7 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	1	2	2	1	2
2	2	2	2	2	2	3
3	2	1	2	2	1	2
4	2	1	2	2	2	2
5	2	1	2	2	2	1
6	2	2	2	2	2	1
7	1	1	2	1	2	1
8	2	2	2	2	1	2
9	2	2	2	2	2	1
10	2	2	2	2	2	1
11	1	1	1	1	1	1
12	2	2	2	2	2	2
13	2	2	2	2	2	2
14	2	2	2	1	2	1
15	2	2	2	2	2	2
16	2	1	2	2	2	2
17	2	1	2	1	1	2
18	2	1	2	2	2	2
19	2	2	2	2	2	2
20	2	2	2	2	2	2
21	2	2	2	2	2	2
22	2	2	2	2	2	2
23	1	1	1	1	1	1
24	1	1	1	1	2	1
25	2	2	2	2	2	1
26	2	2	2	1	2	2
27	1	1	1	1	1	1
28	1	1	1	1	2	1
29	2	1	2	2	2	2
30	2	1	1	1	2	1
31	2	1	2	1	2	2
32	1	1	1	1	1	1
33	2	1	2	2	2	2
34	2	1	2	1	2	2

Intraclass correlation - .8524

Pairwise Comparison - 0.72

Table 7.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 7 B2 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.7.15			1	M.7.15			2	M.7.15			2	M.7.15			1	M.7.15			2	M.7.15		
2	2	M.7.9			2	M.7.9			2	M.7.9			2	M.7.9			3	M.7.9	MHM1					
3	2	M.7.3			1	M.7.3			2	M.7.3			2	M.7.3			2	M.7.3			2	M.7.3		
4	2	M.7.12			1	M.7.12			2	M.7.12			2	M.7.12			2	M.7.12			2	M.7.12		
5	2	M.7.24			1	M.7.24			2	M.7.24			2	M.7.24			1	M.7.24						
6	2	M.7.6			2	M.7.4			2	M.7.4			2	M.7.6			2	M.7.6			1	M.7.6		
7	1	M.7.7			1	M.7.8			2	M.7.7			1	M.7.7			2	M.7.7			1	M.7.7		
8	2	M.7.16			2	M.7.16			2	M.7.16			2	M.7.16			1	M.7.16			2	M.7.16		
9	2	M.7.1			2	M.7.1			2	M.7.1			2	M.7.1			1	M.7.1						
10	2	M.7.8			2	M.7.7			2	M.7.8			2	M.7.8			1	M.7.8						
11	1	M.7.23			1	M.7.23			1	M.7.23			1	M.7.23			1	M.7.23						
12	2	M.7.11			2	M.7.11			2	M.7.11			2	M.7.11			2	M.7.11			2	M.7.11		
13	2	M.7.9			2	M.7.6			2	M.7.9			2	M.7.9			2	M.7.9			2	M.7.9		
14	2	M.7.4			2	M.7.4			2	M.7.4			1	M.7.4			2	M.7.4			1	M.7.4		
15	2	M.7.22			2	M.7.22			2	M.7.22			2	M.7.22			2	M.7.22			2	M.7.22		
16	2	M.7.15			1	M.7.15			2	M.7.15			2	M.7.15			2	M.7.15			2	M.7.15		
17	2	M.7.18			1	M.7.18			2	M.7.18			1	M.7.18			2	M.7.18			2	M.7.18		
18	2	M.7.3			1	M.7.3			2	M.7.3			2	M.7.3			2	M.7.3			2	M.7.3		
19	2	M.7.19			2	M.7.19			2	M.7.19			2	M.7.19			2	M.7.19			2	M.7.19		
20	2	M.7.6			2	M.7.6			2	M.7.4			2	M.7.6			2	M.7.6			2	M.7.6		
21	2	M.7.9			2	M.7.3			2	M.7.9			2	M.7.9			2	M.7.9			2	M.7.9		
22	2	M.7.12			2	M.7.12			2	M.7.12			2	M.7.12			2	M.7.12			2	M.7.12		
23	1	M.7.5			1	M.7.5			1	M.7.5			1	M.7.5			1	M.7.5			1	M.7.5		
24	1	M.7.7			1	M.7.7			1	M.7.7			1	M.7.7			2	M.7.7			1	M.7.7		
25	2	M.7.23			2	M.7.23			2	M.7.23			2	M.7.23			2	M.7.23			1	M.7.23		
26	2	M.7.13			2	M.7.13			2	M.7.13			1	M.7.13			2	M.7.13			2	M.7.13		
27	1	M.7.14			1	M.7.14			1	M.7.14			1	M.7.14			1	M.7.14			1	M.7.14		
28	1	M.7.7			1	M.7.7			1	M.7.7			1	M.7.7			2	M.7.7			1	M.7.7		
29	2	M.7.20			1	M.7.20			2	M.7.20			2	M.7.20			2	M.7.20			2	M.7.19		
30	2	M.7.1			1	M.7.1			1	M.7.1			2	M.7.1			1	M.7.1			1	M.7.1		
31	2	M.7.8			1	M.7.8			2	M.7.8			1	M.7.8			2	M.7.8			2	M.7.8		
32	1	M.7.20			1	M.7.19			1	M.7.20			1	M.7.20			1	M.7.20			1	M.7.20		
33	2	M.7.4			1	M.7.4			2	M.7.4			2	M.7.4			2	M.7.4			2	M.7.4		
34	2	M.7.10			1	M.7.10			2	M.7.10			1	M.7.10			2	M.7.10			2	M.7.10		
Objective Pairwise Comparison: 0.91																								
Standard Pairwise Comparison: 0.97																								

Table 7.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV MATH 2019 Grade 7 B2

	Low	Medium		High
	0	7.2		12
RP.NS.M.7				
RP.NS.M.7A				
M.7.1	9(6)	30(6)		
M.7.2				
M.7.3	3(6)	18(6)	21(1)	
RP.NS.M.7B				
M.7.4	14(12)	6(2)	33(12)	20(1)
M.7.5	23(6)			
M.7.6	6(4)	13(1)	20(5)	
EE.M.7				
EE.M.7A				
M.7.7	10(1)	7(5)	24(6)	28(6)
M.7.8	7(1)	10(5)	31(6)	
EE.M.7B				
M.7.9	13(5)	2(6)	21(5)	
M.7.10	34(12)			
G.M.7				
G.M.7.A				
M.7.11	12(6)			
M.7.12	4(6)	22(12)		
M.7.13	26(6)			
G.M.7.B				

M.7.14	27(6)			
M.7.15	16(6)	1(6)		
M.7.16	8(6)			
SP.M.7				
SP.M.7.A				
M.7.17				
M.7.18	17(6)			
SP.M.7.B				
M.7.19	19(6)	29(1)	32(1)	
M.7.20	32(5)	29(5)		
M.7.21				
M.7.22	15(6)			
SP.M.7.C				
M.7.23	25(6)	11(6)		
M.7.24	5(6)			
M.7.25				
M.7.26				
MHM				
MHM1	2(1)			
MHM2				
MHM3				
MHM4				
MHM5				
MHM6				
MHM7				
MHM8				

Table 7.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 WV MATH 2019 Grade 7 B2

Low		Medium		High	
2.4		7.2		12	
1	10199-2283	M.7.15:6			
2	10199-2512	M.7.9:6		MHM1:1	
3	10199-2160	M.7.3:6			
4	REP10199-2662	M.7.12:6			
5	10199-2454	M.7.24:6			
6	10199-2327	M.7.4:2		M.7.6:4	
7	REP10199-2428	M.7.7:5		M.7.8:1	
8	10199-2361	M.7.16:6			
9	10199-2334	M.7.1:6			
10	REP10199-12647	M.7.7:1		M.7.8:5	
11	10199-2447	M.7.23:6			
12	10199-3014	M.7.11:6			
13	10199-2294	M.7.6:1		M.7.9:5	
14	10199-2582	M.7.4:12			
15	10199-2631	M.7.22:6			
16	10199-2258	M.7.15:6			
17	REP10199-2605	M.7.18:6			
18	10199-2369	M.7.3:6			
19	REP10199-2697	M.7.19:6			
20	10199-2228	M.7.4:1		M.7.6:5	
21	10199-2511	M.7.3:1		M.7.9:5	
22	REP10199-1141	M.7.12:12			
23	10199-2280	M.7.5:6			
24	REP10199-2293	M.7.7:6			
25	10199-2446	M.7.23:6			
26	10199-3008	M.7.13:6			
27	10199-11418	M.7.14:6			
28	10199-2434	M.7.7:6			
29	10199-2087	M.7.19:1		M.7.20:5	
30	10199-2310	M.7.1:6			
31	REP10199-2550	M.7.8:6			
32	10199-2074	M.7.19:1		M.7.20:5	
33	REP10199-2530	M.7.4:12			
34	10199-2320	M.7.10:12			

Table 7.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV MATH 2019 Grade 7 B2

Low DOK		Matched DOK		High DOK

RP.NS.M.7				
RP.NS.M.7A				
M.7.1: [2]	9:(6)[2]	30:(6)[1]		
M.7.2				
M.7.3: [2]	3:(6)[2]	18:(6)[2]	21:(1)[2]	
RP.NS.M.7B				
M.7.4: [2]	6:(2)[2]	14:(12)[2]	20:(1)[2]	33:(12)[2]
M.7.5: [2]	23:(6)[1]			
M.7.6: [2]	6:(4)[2]	13:(1)[2]	20:(5)[2]	
EE.M.7				
EE.M.7A				
M.7.7: [1]	7:(5)[1]	10:(1)[2]	24:(6)[1]	28:(6)[1]
M.7.8: [2]	7:(1)[1]	10:(5)[2]	31:(6)[2]	
EE.M.7B				
M.7.9: [2]	2:(6)[2]	13:(5)[2]	21:(5)[2]	
M.7.10: [2]	34:(12)[2]			
G.M.7				
G.M.7.A				
M.7.11: [2]	12:(6)[2]			
M.7.12: [2]	4:(6)[2]	22:(12)[2]		
M.7.13: [2]	26:(6)[2]			
G.M.7.B				
M.7.14: [2]	27:(6)[1]			
M.7.15: [2]	1:(6)[2]	16:(6)[2]		
M.7.16: [2]	8:(6)[2]			
SP.M.7				
SP.M.7.A				
M.7.17				

M.7.18: [3]	17:(6)[2]			
SP.M.7.B				
M.7.19: [2]	19:(6)[2]	29:(1)[2]	32:(1)[1]	
M.7.20: [2]	29:(5)[2]	32:(5)[1]		
M.7.21				
M.7.22: [2]	15:(6)[2]			
SP.M.7.C				
M.7.23: [1]	11:(6)[1]	25:(6)[2]		
M.7.24: [2]	5:(6)[2]			
M.7.25				
M.7.26				
MHM				
MHM1: [3]	2:(1)[3]			
MHM2				
MHM3				
MHM4				
MHM5				
MHM6				
MHM7				
MHM8				

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 7B2 MATH

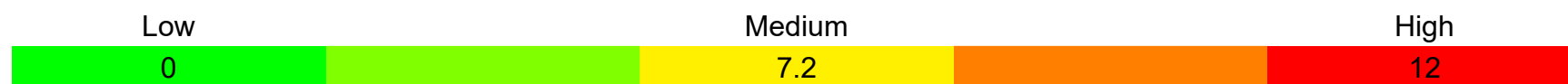
[illegible]

Table 7.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 7B2 MATH

Low	Medium	High
2.4	7.2	12

1 10199-2283	Exact:6	
2 10199-2512	Exact:6	
3 10199-2160	Exact:6	
4 REP10199-2662	Exact:6	
5 10199-2454	Exact:5	Partial:1
6 10199-2327	Exact:4	Partial:2
7 REP10199-2428	Exact:5	Partial:1
8 10199-2361	Exact:5	Partial:1
9 10199-2334	Exact:6	
10 REP10199-12647	Exact:6	
11 10199-2447	Exact:6	
12 10199-3014	Exact:6	
13 10199-2294	Exact:6	
14 10199-2582	Exact:10	Partial:2
15 10199-2631	Exact:6	
16 10199-2258	Exact:6	
17 REP10199-2605	Exact:6	
18 10199-2369	Exact:6	
19 REP10199-2697	Exact:1	Partial:5
20 10199-2228	Exact:5	Partial:1
21 10199-2511	Exact:4	Partial:2
22 REP10199-1141	Exact:12	
23 10199-2280	Exact:5	Partial:1
24 REP10199-2293	Exact:6	
25 10199-2446	Exact:6	
26 10199-3008	Exact:6	
27 10199-11418	Exact:5	Partial:1
28 10199-2434	Exact:6	
29 10199-2087	Exact:2	Partial:4
30 10199-2310	Exact:6	
31 REP10199-2550	Exact:6	
32 10199-2074	Exact:4	Partial:2
33 REP10199-2530	Exact:10	Partial:2
34 10199-2320	Exact:10	Partial:2

Batch 4 West Virginia Math Grade 7

Table 7.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 7 B4 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	2	6	2	6	100	8	4	YES
EE.M.7 Expressions and Equations	2	4	1 2	1 3	25 75	5.25	3.5	NO
G.M.7 Geometry	2	6	2	6	100	7.75	4.5	YES
SP.M.7 Statistics and Probability	3	10	1 2 3	1 6 3	10 60 30	6.75	4.5	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	17	34	1 2 3	2 23 9	6 68 26	27.75	16.5	

Table 7.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 7 B4 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	2	6	8	4	27.5	15	70	14	2.5	5	YES
EE.M.7 Expressions and Equations	2	4	5.25	3.5	4.76	8	71.43	25	23.81	16	YES
G.M.7 Geometry	2	6	7.75	4.5	5	6	65	44	30	47	YES
SP.M.7 Statistics and Probability	3	10	6.75	4.5	33.33	11	59.26	17	7.41	6	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	17	34	27.75	16.5	17.12	9.5	72.97	22.5	9.91	13.1	

Table 7.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 7 B4 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	2	6	8	4	3.75	1.89	62.5	31.55	YES	39	19	0.86	0.09	YES
EE.M.7 Expressions and Equations	2	4	5.25	3.5	2.25	1.5	56.25	37.5	YES	15	10	0.9	0.08	YES
G.M.7 Geometry	2	6	7.75	4.5	4	2	66.67	33.33	YES	26	5	0.77	0.15	YES
SP.M.7 Statistics and Probability	3	10	6.75	4.5	5.5	3.7	55	36.97	YES	20	13	0.9	0.07	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	17	34	27.75	16.5	3.1	2.08	48.08	27		20	14	0.86	0.08	

Table 7.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 7 B4 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
RP.NS.M.7 Ratios and Proportional Relationships and the Number System	YES	YES	YES	YES
EE.M.7 Expressions and Equations	NO	YES	YES	YES
G.M.7 Geometry	YES	YES	YES	YES
SP.M.7 Statistics and Probability	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 7.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 7 B4

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	3	3	3		
2	2	2	2	1		
3	2	2	2	2		
4	2	2	1			
5	2	2	2			
6	2	2	2			
7	1	2	2			
8	2	2	2			
9	1	1	2			
10	2	2	2			
11	2	2	2			
12	2	1	2			
13	2	2	2			
14	2	2	2			
15	2	2	2			
16	2	1	2			
17	2	2	2			
18	2	1	2			
19	2	2	2			
20	2	2	2			
21	2	2	2			
22	2	3	2			
23	1	2	1			
24	1	1	2			
25	1	1	1			
26	1	1	2			
27	2	2	2			
28	2	2	2			
29	2	2	2			
30	2	2	2			
31	2	2	2			
32	2	1	2			
33	2	3	2			
34	2	2	2			

Intraclass correlation - .7297

Pairwise Comparison - 0.72

Table 7.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 7 B4 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.7.14			3	M.7.14			3	M.7.16			3	M.7.14		
2	2	M.7.3			2	M.7.3			2	M.7.3			1	M.7.3		
3	2	M.7.3			2	M.7.3			2	M.7.3			2	M.7.3		
4	2	M.7.23			2	M.7.23			1	M.7.23						
5	2	M.7.11			2	M.7.11			2	M.7.11						
6	2	M.7.6			2	M.7.6			2	M.7.4						
7	1	M.7.7			2	M.7.7			2	M.7.8						
8	2	M.7.12			2	M.7.12			2	M.7.12						
9	1	M.7.15			1	M.7.15			2	M.7.15						
10	2	M.7.3			2	M.7.3			2	M.7.3						
11	2	M.7.24			2	M.7.24			2	M.7.24						
12	2	M.7.8			1	M.7.8			2	M.7.8						
13	2	M.7.6			2	M.7.6			2	M.7.4						
14	2	M.7.22			2	M.7.22			2	M.7.20						
15	2	M.7.9			2	M.7.9			2	M.7.9						
16	2	M.7.18			1	M.7.18			2	M.7.18						
17	2	M.7.15			2	M.7.15			2	M.7.15						
18	2	M.7.1			1	M.7.1			2	M.7.1						
19	2	M.7.12			2	M.7.12			2	M.7.12						
20	2	M.7.24			2	M.7.24			2	M.7.24						
21	2	M.7.22			2	M.7.22			2	M.7.22						
22	2	M.7.9			3	M.7.9			2	M.7.9						
23	1	M.7.1			2	M.7.1			1	M.7.1						
24	1	M.7.7			1	M.7.7			2	M.7.9						
25	1	M.7.5			1	M.7.5			1	M.7.5						
26	1	M.7.19			1	M.7.19			2	M.7.19						
27	2	M.7.12			2	M.7.12			2	M.7.12						
28	2	M.7.8			2	M.7.8			2	M.7.8						
29	2	M.7.13			2	M.7.13			2	M.7.13						
30	2	M.7.1			2	M.7.1			2	M.7.1						
31	2	M.7.7			2	M.7.7			2	M.7.7						
32	2	M.7.21			1	M.7.21			2	M.7.21						
33	2	M.7.4			3	M.7.4			2	M.7.4						
34	2	M.7.26			2	M.7.26			2	M.7.26						
Objective Pairwise Comparison: 0.89																
Standard Pairwise Comparison: 1																

Table 7.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV MATH 2019 Grade 7 B4

Low	Medium		High
0	3.6		6
RP.NS.M.7			
RP.NS.M.7A			
M.7.1	18(3)	23(3)	30(3)
M.7.2			
M.7.3	2(4)	3(4)	10(3)
RP.NS.M.7B			
M.7.4	13(1)	6(1)	33(3)
M.7.5	25(3)		
M.7.6	6(2)	13(2)	
EE.M.7			
EE.M.7A			
M.7.7	7(2)	24(2)	31(3)
M.7.8	28(3)	7(1)	12(3)
EE.M.7B			
M.7.9	22(3)	24(1)	15(3)
M.7.10			
G.M.7			
G.M.7.A			
M.7.11	5(3)		
M.7.12	8(6)	19(3)	27(3)
M.7.13	29(3)		
G.M.7.B			

M.7.14	1(3)		
M.7.15	9(3)	17(6)	
M.7.16	1(1)		
SP.M.7			
SP.M.7.A			
M.7.17			
M.7.18	16(3)		
SP.M.7.B			
M.7.19	26(3)		
M.7.20	14(1)		
M.7.21	32(3)		
M.7.22	14(2)	21(3)	
SP.M.7.C			
M.7.23	4(3)		
M.7.24	11(3)	20(3)	
M.7.25			
M.7.26	34(3)		
MHM			
MHM1			
MHM2			
MHM3			
MHM4			
MHM5			
MHM6			
MHM7			
MHM8			

Table 7.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV MATH 2019 Grade 7 B4

Low		Medium		High	
1.2		3.6		6	
1 10199-2174				M.7.14:3	M.7.16:1
2 REP10199-2166				M.7.3:4	
3 10199-2473				M.7.3:4	
4 10199-2588				M.7.23:3	
5 10199-2202				M.7.11:3	
6 10199-2226				M.7.4:1	
7 10199-2436				M.7.7:2	M.7.8:1
8 REP10199-1141				M.7.12:6	
9 10199-2253				M.7.15:3	
10 10199-2368				M.7.3:3	
11 REP10199-2448				M.7.24:3	
12 REP10199-2550				M.7.8:3	
13 10199-2328				M.7.4:1	M.7.6:2
14 10199-2565				M.7.20:1	M.7.22:2
15 10199-2514				M.7.9:3	
16 REP10199-2605				M.7.18:3	
17 10199-9725				M.7.15:6	
18 10199-2326				M.7.1:3	
19 10199-2665				M.7.12:3	
20 10199-2461				M.7.24:3	
21 10199-2570				M.7.22:3	
22 10199-2515				M.7.9:3	
23 10199-2225				M.7.1:3	
24 10199-10023				M.7.7:2	
25 10199-2142				M.7.5:3	
26 10199-2119				M.7.19:3	
27 10199-3018				M.7.12:3	
28 REP10199-2175				M.7.8:3	
29 10199-13404				M.7.13:3	
30 10199-2481				M.7.1:3	
31 10199-2437				M.7.7:3	
32 10199-2482				M.7.21:3	
33 10199-2536				M.7.4:3	
34 10199-2625				M.7.26:3	

Table 7.9

Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])

WV MATH 2019 Grade 7 B4

Low DOK		Matched DOK		High DOK

RP.NS.M.7			
RP.NS.M.7A			
M.7.1: [2]	18:(3)[2]	23:(3)[1]	30:(3)[2]
M.7.2			
M.7.3: [2]	2:(4)[2]	3:(4)[2]	10:(3)[2]
RP.NS.M.7B			
M.7.4: [2]	6:(1)[2]	13:(1)[2]	33:(3)[2]
M.7.5: [2]	25:(3)[1]		
M.7.6: [2]	6:(2)[2]	13:(2)[2]	
EE.M.7			
EE.M.7A			
M.7.7: [1]	7:(2)[2]	24:(2)[1]	31:(3)[2]
M.7.8: [2]	7:(1)[2]	12:(3)[2]	28:(3)[2]
EE.M.7B			
M.7.9: [2]	15:(3)[2]	22:(3)[2]	24:(1)[2]
M.7.10			
G.M.7			
G.M.7.A			
M.7.11: [2]	5:(3)[2]		
M.7.12: [2]	8:(6)[2]	19:(3)[2]	27:(3)[2]
M.7.13: [2]	29:(3)[2]		
G.M.7.B			
M.7.14: [2]	1:(3)[3]		
M.7.15: [2]	9:(3)[1]	17:(6)[2]	
M.7.16: [2]	1:(1)[3]		
SP.M.7			
SP.M.7.A			
M.7.17			

M.7.18: [3]	16:(3)[2]		
SP.M.7.B			
M.7.19: [2]	26:(3)[1]		
M.7.20: [2]	14:(1)[2]		
M.7.21: [2]	32:(3)[2]		
M.7.22: [2]	14:(2)[2]	21:(3)[2]	
SP.M.7.C			
M.7.23: [1]	4:(3)[2]		
M.7.24: [2]	11:(3)[2]	20:(3)[2]	
M.7.25			
M.7.26: [3]	34:(3)[2]		
MHM			
MHM1			
MHM2			
MHM3			
MHM4			
MHM5			
MHM6			
MHM7			
MHM8			

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 7B4 MATH

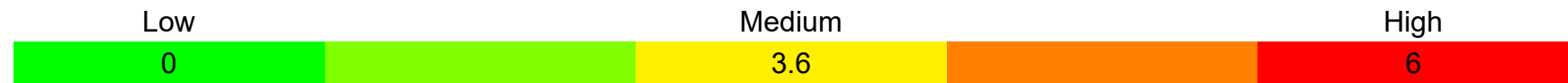
[illegible]

Table 7.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 7B4 MATH

Low		Medium		High	
1.2		3.6		6	
1	10199-2174	Exact:2	Partial:1		
2	REP10199-2166	Exact:3			
3	10199-2473	Exact:1	Partial:1	Minimal:1	
4	10199-2588	Exact:3			
5	10199-2202	Exact:3			
6	10199-2226	Exact:2	Partial:1		
7	10199-2436	Exact:2	Partial:1		
8	REP10199-1141	Exact:6			
9	10199-2253	Exact:3			
10	10199-2368	Exact:3			
11	REP10199-2448	Exact:3			
12	REP10199-2550	Exact:3			
13	10199-2328	Exact:2	Partial:1		
14	10199-2565	Exact:2	Partial:1		
15	10199-2514	Exact:3			
16	REP10199-2605	Exact:3			
17	10199-9725	Exact:6			
18	10199-2326	Exact:3			
19	10199-2665	Exact:3			
20	10199-2461	Exact:3			
21	10199-2570	Exact:3			
22	10199-2515	Exact:2	Partial:1		
23	10199-2225	Exact:3			
24	10199-10023	Exact:2	Partial:1		
25	10199-2142	Exact:3			
26	10199-2119	Exact:3			
27	10199-3018	Exact:3			
28	REP10199-2175	Exact:3			
29	10199-13404	Exact:3			
30	10199-2481	Exact:3			
31	10199-2437	Exact:3			
32	10199-2482	Exact:3			
33	10199-2536	Exact:3			
34	10199-2625	Exact:3			

Batch 1 West Virginia Math Grade 8

Table 8.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 8 B1 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	4	10	1 2	4 6	40 60	12	0	YES
F.M.8 Functions (F)	2	5	1 2	1 4	20 80	9	0	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	4	13	1 2	1 12	7.69 92.31	15	0	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	1	0.63	NO
Total	18	36	1 2 3	6 24 6	17 67 17	37	0.63	

Table 8.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 8 B1 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	4	10	12	0	30.56	7	69.44	7	0	0	YES
F.M.8 Functions (F)	2	5	9	0	7.41	9	88.89	17	3.7	9	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	4	13	15	0	27.78	15	65.56	15	6.67	0	YES
MHM Mathematical Habits of Mind	8	8	1	0.63	100	0	0	0	0	0	NO
Total	18	36	37	0.63	25.68	7.9	70.72	9.6	3.6	2.3	

Table 8.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 8 B1 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	4	10	12	0	9	0	90	0	YES	31	1	0.83	0	YES
F.M.8 Functions (F)	2	5	9	0	5	0	100	0	YES	26	0	0.82	0	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	4	13	15	0	11.5	0.55	88.46	4.21	YES	40	1	0.82	0.01	YES
MHM Mathematical Habits of Mind	8	8	1	0.63	0.83	0.41	10.42	5.1	NO	3	2	1	0	YES
Total	18	36	37	0.63	6.6	4.68	72.22	42		25	16	0.87	0.09	

Table 8.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 8 B1 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	YES	YES	YES	YES
F.M.8 Functions (F)	YES	YES	YES	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NO	NO	YES

Table 8.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 8 B1

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	1	1	2	1	2	1
2	2	1	2	2	1	2
3	2	1	2	2	2	2
4	2	2	2	2	2	3
5	2	2	2	2	1	1
6	1	1	1	1	2	1
7	1	1	1	1	1	1
8	1	1	1	1	2	1
9	2	1	2	2	2	1
10	1	2	2	2	2	2
11	2	2	2	2	2	2
12	2	1	2	2	2	1
13	2	2	2	2	1	2
14	1	1	1	1	1	1
15	2	2	2	2	2	2
16	1	1	1	2	1	1
17	1	1	1	1	1	1
18	2	2	2	2	2	2
19	2	1	2	2	2	1
20	2	2	2	2	2	2
21	1	1	1	1	1	1
22	1	1	1	1	1	1
23	2	2	2	1	2	1
24	2	2	2	2	2	2
25	2	1	2	2	2	1
26	2	2	2	2	1	2
27	1	1	1	1	1	2
28	2	2	2	2	2	2
29	1	1	1	1	1	1
30	2	2	2	2	2	2
31	2	2	2	2	2	2
32	2	2	2	2	2	2
33	1	2	2	2	2	2
34	2	2	2	2	2	1

Intraclass correlation - .8981

Pairwise Comparison - 0.77

Table 8.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 8 B1 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	1	M.8.26			1	M.8.26			2	M.8.26			1	M.8.26			2	M.8.26			1	M.8.26		
2	2	M.8.20			1	M.8.20			2	M.8.20			2	M.8.20			1	M.8.20			2	M.8.20		
3	2	M.8.15			1	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15		
4	2	M.8.12			2	M.8.12			2	M.8.12			2	M.8.12			2	M.8.12			3	M.8.12		
5	2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14			1	M.8.14			1	M.8.14		
6	1	M.8.2			1	M.8.2			1	M.8.2			1	M.8.2			2	M.8.2			1	M.8.2		
7	1	M.8.11			1	M.8.11			1	M.8.11			1	M.8.11			1	M.8.11			1	M.8.11		
8	1	M.8.17			1	M.8.17			1	M.8.17			1	M.8.17			2	M.8.17			1	M.8.17		
9	2	M.8.28			1	M.8.28			2	M.8.21			2	M.8.28			2	M.8.28			1	M.8.28		
10	1	M.8.8			2	M.8.8			2	M.8.8			2	M.8.8			2	M.8.8			2	M.8.8		
11	2	M.8.24			2	M.8.24			2	M.8.24			2	M.8.24			2	M.8.24			2	M.8.24		
12	2	M.8.7			1	M.8.7			2	M.8.7			2	M.8.7			2	M.8.7			1	M.8.7		
13	2	M.8.19			2	M.8.19			2	M.8.19			2	M.8.19			1	M.8.19			2	M.8.19		
14	1	M.8.4			1	M.8.4			1	M.8.4			1	M.8.4			1	M.8.4			1	M.8.4		
15	2	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15		
16	1	M.8.22			1	M.8.22			1	M.8.22			2	M.8.22			1	M.8.22			1	M.8.22		
17	1	M.8.3			1	M.8.3			1	M.8.3			1	M.8.3			1	M.8.3			1	M.8.3		
18	2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14		
19	2	M.8.28			1	M.8.28			2	M.8.28			2	M.8.28			2	M.8.28			1	M.8.28		
20	2	MHM1	M.8.22		2	M.8.21	MHM1		2	M.8.22	MHM1		2	M.8.22	MHM1		2	M.8.21	MHM1		2	M.8.22		
21	1	M.8.7			1	M.8.7			1	M.8.7			1	M.8.7			1	M.8.7			1	M.8.7		
22	1	M.8.1			1	M.8.1			1	M.8.1			1	M.8.1			1	M.8.1			1	M.8.1		
23	2	M.8.25			2	M.8.25			2	M.8.25			1	M.8.25			2	M.8.25			1	M.8.25		
24	2	M.8.13			2	M.8.13			2	M.8.13			2	M.8.13			2	M.8.13			2	M.8.13		
25	2	M.8.18			1	M.8.18			2	M.8.18			2	M.8.18			2	M.8.18			1	M.8.18		
26	2	M.8.5			2	M.8.5			2	M.8.5			2	M.8.5			1	M.8.5			2	M.8.5		
27	1	M.8.11			1	M.8.11			1	M.8.11	MHM1		1	M.8.11			1	M.8.11			2	M.8.11		
28	2	M.8.23			2	M.8.23			2	M.8.23			2	M.8.23			2	M.8.23			2	M.8.23		
29	1	M.8.5			1	M.8.5			1	M.8.5			1	M.8.5			1	M.8.5			1	M.8.5		
30	2	M.8.25			2	M.8.25			2	M.8.25			2	M.8.25			2	M.8.25			2	M.8.25		
31	2	M.8.10			2	M.8.10			2	M.8.10			2	M.8.10			2	M.8.10			2	M.8.10		
32	2	M.8.16			2	M.8.16			2	M.8.16			2	M.8.16			2	M.8.16			2	M.8.16		
33	1	M.8.9			2	M.8.9			2	M.8.9			2	M.8.9			2	M.8.9			2	M.8.9		
34	2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14			1	M.8.14		
Objective Pairwise Comparison: 0.96																								
Standard Pairwise Comparison: 0.98																								

Table 8.7 *Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)*
WV MATH 2019 Grade 8 B1

Low 0		Medium 7.2		High 12
NS.EE.M.8				
NS.EE.M.8.A				
M.8.1	22(6)			
M.8.2	6(6)			
NS.EE.M.8.B				
M.8.3	17(6)			
M.8.4	14(12)			
M.8.5	26(6)	29(6)		
M.8.6				
NS.EE.M.8.C				
M.8.7	12(6)	21(6)		
M.8.8	10(6)			
NS.EE.M.8.D				
M.8.9	33(6)			
M.8.10	31(6)			
F.M.8				
F.M.8.A				
M.8.11	7(6)	27(6)		
M.8.12	4(6)			
M.8.13	24(6)			
F.M.8.B				
M.8.14	18(6)	5(6)	34(6)	
M.8.15	15(6)	3(6)		

G.SP.M.8			
G.SP.M.8.A			
M.8.16	32(6)		
M.8.17	8(6)		
M.8.18	25(6)		
M.8.19	13(6)		
M.8.20	2(6)		
G.SP.M.8.B			
M.8.21	9(1)	20(2)	
M.8.22	16(6)	20(4)	
M.8.23	28(6)		
G.SP.M.8.C			
M.8.24	11(6)		
G.SP.M.8.D			
M.8.25	30(12)	23(6)	
M.8.26	1(6)		
M.8.27			
M.8.28	9(5)	19(6)	
MHM			
MHM1	27(1)	20(5)	
MHM2			
MHM3			
MHM4			
MHM5			
MHM6			
MHM7			
MHM8			

Table 8.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV MATH 2019 Grade 8 B1

Low		Medium		High	
2.4		7.2		12	
1	10199-2892	M.8.26:6			
2	10199-2758	M.8.20:6			
3	REP10199-2861	M.8.15:6			
4	10199-2889	M.8.12:6			
5	REP10199-2780	M.8.14:6			
6	10199-2947	M.8.2:6			
7	10199-2723	M.8.11:6			
8	REP10199-3006	M.8.17:6			
9	REP10199-2521	M.8.21:1		M.8.28:5	
10	10199-2799	M.8.8:6			
11	10199-2362	M.8.24:6			
12	10199-2796	M.8.7:6			
13	10199-3918	M.8.19:6			
14	10199-2773	M.8.4:12			
15	10199-2737	M.8.15:6			
16	10199-3736	M.8.22:6			
17	10199-2703	M.8.3:6			
18	REP10199-2768	M.8.14:6			
19	10199-2518	M.8.28:6			
20	10199-4176	M.8.21:2		M.8.22:4	MHM1:5
21	10199-2795	M.8.7:6			
22	10199-2589	M.8.1:6			
23	REP10199-2708	M.8.25:6			
24	10199-2698	M.8.13:6			
25	10199-2852	M.8.18:6			
26	10199-2843	M.8.5:6			
27	REP10199-2743	M.8.11:6		MHM1:1	
28	10199-2815	M.8.23:6			
29	10199-2765	M.8.5:6			
30	10199-2827	M.8.25:12			
31	10199-2939	M.8.10:6			
32	10199-11424	M.8.16:6			
33	10199-2901	M.8.9:6			
34	REP10199-2781	M.8.14:6			

Table 8.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
 WV MATH 2019 Grade 8 B1

Low DOK		Matched DOK		High DOK

NS.EE.M.8			
NS.EE.M.8.A			
M.8.1: [1]	22:(6)[1]		
M.8.2: [2]	6:(6)[1]		
NS.EE.M.8.B			
M.8.3: [1]	17:(6)[1]		
M.8.4: [1]	14:(12)[1]		
M.8.5: [2]	26:(6)[2]	29:(6)[1]	
M.8.6			
NS.EE.M.8.C			
M.8.7: [2]	12:(6)[2]	21:(6)[1]	
M.8.8: [2]	10:(6)[2]		
NS.EE.M.8.D			
M.8.9: [2]	33:(6)[2]		
M.8.10: [2]	31:(6)[2]		
F.M.8			
F.M.8.A			
M.8.11: [1]	7:(6)[1]	27:(6)[1]	
M.8.12: [2]	4:(6)[2]		
M.8.13: [2]	24:(6)[2]		
F.M.8.B			
M.8.14: [2]	5:(6)[2]	18:(6)[2]	34:(6)[2]
M.8.15: [2]	3:(6)[2]	15:(6)[2]	
G.SP.M.8			
G.SP.M.8.A			
M.8.16: [2]	32:(6)[2]		
M.8.17: [2]	8:(6)[1]		
M.8.18: [2]	25:(6)[2]		
M.8.19: [2]	13:(6)[2]		

M.8.20: [2]	2:(6)[2]		
G.SP.M.8.B			
M.8.21: [2]	9:(1)[2]	20:(2)[2]	
M.8.22: [2]	16:(6)[1]	20:(4)[2]	
M.8.23: [1]	28:(6)[2]		
G.SP.M.8.C			
M.8.24: [2]	11:(6)[2]		
G.SP.M.8.D			
M.8.25: [2]	23:(6)[2]	30:(12)[2]	
M.8.26: [2]	1:(6)[1]		
M.8.27			
M.8.28: [2]	9:(5)[2]	19:(6)[2]	
MHM			
MHM1: [3]	20:(5)[2]	27:(1)[1]	
MHM2			
MHM3			
MHM4			
MHM5			
MHM6			
MHM7			
MHM8			

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 8B1 MATH

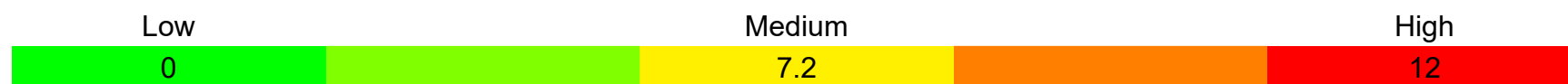
[illegible]

Table 8.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 8B1 MATH

Low		Medium		High	
2.4		7.2		12	
1	10199-2892	Exact:6			
2	10199-2758	Exact:5		Partial:1	
3	REP10199-2861	Exact:6			
4	10199-2889	Exact:6			
5	REP10199-2780	Exact:6			
6	10199-2947	Exact:6			
7	10199-2723	Exact:6			
8	REP10199-3006	Exact:6			
9	REP10199-2521	Exact:6			
10	10199-2799	Exact:5		Partial:1	
11	10199-2362	Exact:5		Partial:1	
12	10199-2796	Exact:5		Partial:1	
13	10199-3918	Exact:6			
14	10199-2773	Exact:10		Partial:2	
15	10199-2737	Exact:5		Partial:1	
16	10199-3736	Exact:6			
17	10199-2703	Exact:6			
18	REP10199-2768	Partial:5		Minimal:1	
19	10199-2518	Exact:5		Partial:1	
20	10199-4176	Partial:6			
21	10199-2795	Exact:5		Partial:1	
22	10199-2589	Exact:6			
23	REP10199-2708	Exact:6			
24	10199-2698	Exact:6			
25	10199-2852	Exact:6			
26	10199-2843	Exact:6			
27	REP10199-2743	Exact:6			
28	10199-2815	Exact:6			
29	10199-2765	Exact:6			
30	10199-2827	Exact:12			
31	10199-2939	Exact:5		Partial:1	
32	10199-11424	Exact:6			
33	10199-2901	Exact:5		Partial:1	
34	REP10199-2781	Exact:5		Partial:1	

Batch 2 West Virginia Math Grade 8

Table 8.1 *Categorical Concurrence between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 8 B2 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	4	10	1 2	4 6	40 60	12.17	0.41	YES
F.M.8 Functions (F)	2	5	1 2	1 4	20 80	9	0	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	4	13	1 2	1 12	7.69 92.31	14.83	0.41	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	18	36	1 2 3	6 24 6	17 67 17	36	0	

Table 8.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Six Reviewers WV MATH 2019 Grade 8 B2 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	4	10	12.17	0.41	20.62	5	69.76	7	9.62	3	YES
F.M.8 Functions (F)	2	5	9	0	27.78	14	68.52	20	3.7	9	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	4	13	14.83	0.41	46.19	17	53.81	17	0	0	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	18	36	36	0	32.87	9.8	62.96	11.7	4.17	2.3	

Table 8.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Six Reviewers WV MATH 2019 Grade 8 B2 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	4	10	12.17	0.41	9.33	0.52	93.33	5.16	YES	33	1	0.84	0.01	YES
F.M.8 Functions (F)	2	5	9	0	4	0	80	0	YES	26	0	0.92	0	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	4	13	14.83	0.41	10.17	0.75	78.21	5.79	YES	41	1	0.78	0.02	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	18	36	36	0	5.9	4.78	62.88	42		25	18	0.85	0.1	

Table 8.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers WV MATH 2019 Grade 8 B2 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	YES	YES	YES	YES
F.M.8 Functions (F)	YES	YES	YES	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 8.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 8 B2

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1	2	2	2	1	2	1
2	2	1	2	2	2	2
3	1	1	2	2	1	2
4	1	1	2	1	1	1
5	2	1	2	2	1	1
6	2	1	2	1	1	1
7	1	1	1	1	1	2
8	2	2	2	2	2	2
9	2	2	2	2	2	2
10	2	2	2	1	2	1
11	1	2	1	1	1	1
12	1	1	2	1	1	1
13	1	1	1	1	1	1
14	2	2	2	2	2	2
15	2	1	2	2	1	1
16	1	2	1	1	1	1
17	1	1	1	1	1	1
18	3	3	2	3	3	3
19	2	2	2	2	2	2
20	2	2	2	2	2	2
21	2	2	2	2	1	1
22	2	2	2	2	2	1
23	1	1	1	1	1	1
24	1	1	2	1	1	1
25	2	2	2	2	1	1
26	1	1	2	1	1	1
27	1	1	1	1	1	2
28	1	1	1	1	1	1
29	1	1	1	1	1	2
30	2	2	2	2	2	2
31	2	2	2	3	2	2
32	1	2	2	2	2	2
33	1	2	2	1	2	1
34	1	1	1	1	2	1

Intraclass correlation - .8948

Pairwise Comparison - 0.7

Table 8.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 8 B2 Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.8.25			2	M.8.25			2	M.8.25			1	M.8.25			2	M.8.25			1	M.8.25		
2	2	M.8.15			1	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15		
3	1	M.8.20			1	M.8.20			2	M.8.20			2	M.8.20			1	M.8.20			2	M.8.20		
4	1	M.8.13			1	M.8.13			2	M.8.13			1	M.8.13			1	M.8.13			1	M.8.13		
5	2	M.8.14			1	M.8.14			2	M.8.14			2	M.8.14			1	M.8.14			1	M.8.14		
6	2	M.8.2			1	M.8.2			2	M.8.2			1	M.8.2			1	M.8.2			1	M.8.2		
7	1	M.8.11			1	M.8.11			1	M.8.11			1	M.8.11			1	M.8.11			2	M.8.11		
8	2	M.8.7			2	M.8.7			2	M.8.7			2	M.8.7			2	M.8.7			2	M.8.7		
9	2	M.8.17			2	M.8.17			2	M.8.17			2	M.8.17			2	M.8.17			2	M.8.17		
10	2	M.8.26			2	M.8.26			2	M.8.26			1	M.8.26			2	M.8.26			1	M.8.26		
11	1	M.8.24			2	M.8.24			1	M.8.24			1	M.8.24			1	M.8.24			1	M.8.24		
12	1	M.8.22			1	M.8.21			2	M.8.21			1	M.8.22			1	M.8.22			1	M.8.22		
13	1	M.8.5			1	M.8.5			1	M.8.5			1	M.8.5			1	M.8.5			1	M.8.5		
14	2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14		
15	2	M.8.28			1	M.8.28			2	M.8.28			2	M.8.28			1	M.8.28			1	M.8.28		
16	1	M.8.19			2	M.8.19			1	M.8.19			1	M.8.19			1	M.8.4			1	M.8.19		
17	1	M.8.4			1	M.8.4			1	M.8.4			1	M.8.4			1	M.8.4			1	M.8.4		
18	3	M.8.8			3	M.8.8			2	M.8.8			3	M.8.8			3	M.8.8			3	M.8.8		
19	2	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15		
20	2	M.8.10			2	M.8.10			2	M.8.10			2	M.8.10			2	M.8.10			2	M.8.10		
21	2	M.8.17			2	M.8.17			2	M.8.17			2	M.8.17			1	M.8.17			1	M.8.17		
22	2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14			2	M.8.14			1	M.8.14		
23	1	M.8.1			1	M.8.1			1	M.8.1			1	M.8.1			1	M.8.1			1	M.8.1		
24	1	M.8.22			1	M.8.22			2	M.8.22			1	M.8.22			1	M.8.22			1	M.8.22		
25	2	M.8.18			2	M.8.18			2	M.8.18			2	M.8.18			1	M.8.18			1	M.8.18		
26	1	M.8.3			1	M.8.3			2	M.8.3			1	M.8.3			1	M.8.3			1	M.8.3		
27	1	M.8.11			1	M.8.11			1	M.8.11			1	M.8.11			1	M.8.11			2	M.8.11		
28	1	M.8.6			1	M.8.5			1	M.8.5			1	M.8.5			1	M.8.6			1	M.8.5		
29	1	M.8.22			1	M.8.22			1	M.8.22			1	M.8.22			1	M.8.22			2	M.8.22		
30	2	M.8.25			2	M.8.25			2	M.8.25			2	M.8.25			2	M.8.25			2	M.8.25		
31	2	M.8.7			2	M.8.7			2	M.8.7			3	M.8.7			2	M.8.7			2	M.8.7		
32	1	M.8.9			2	M.8.9			2	M.8.9			2	M.8.9			2	M.8.9			2	M.8.9		
33	1	M.8.16			2	M.8.16			2	M.8.16			1	M.8.16			2	M.8.16			1	M.8.16		
34	1	M.8.13			1	M.8.13			1	M.8.13			1	M.8.13			2	M.8.13			1	M.8.13		
Objective Pairwise Comparison: 0.96																								
Standard Pairwise Comparison: 0.99																								

Table 8.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV MATH 2019 Grade 8 B2

Low		Medium		High
0		7.2		12
NS.EE.M.8				
NS.EE.M.8.A				
M.8.1	23(6)			
M.8.2	6(6)			
NS.EE.M.8.B				
M.8.3	26(6)			
M.8.4	17(6)	16(1)		
M.8.5	28(4)	13(6)		
M.8.6	28(2)			
NS.EE.M.8.C				
M.8.7	8(6)	31(6)		
M.8.8	18(6)			
NS.EE.M.8.D				
M.8.9	32(6)			
M.8.10	20(12)			
F.M.8				
F.M.8.A				
M.8.11	7(6)	27(6)		
M.8.12				
M.8.13	34(6)	4(6)		
F.M.8.B				
M.8.14	5(6)	14(6)	22(6)	
M.8.15	19(6)	2(6)		

G.SP.M.8			
G.SP.M.8.A			
M.8.16	33(6)		
M.8.17	9(6)	21(6)	
M.8.18	25(6)		
M.8.19	16(5)		
M.8.20	3(6)		
G.SP.M.8.B			
M.8.21	12(2)		
M.8.22	12(4)	24(6)	29(6)
M.8.23			
G.SP.M.8.C			
M.8.24	11(6)		
G.SP.M.8.D			
M.8.25	1(6)	30(12)	
M.8.26	10(6)		
M.8.27			
M.8.28	15(6)		
MHM			
MHM1			
MHM2			
MHM3			
MHM4			
MHM5			
MHM6			
MHM7			
MHM8			

Table 8.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
WV MATH 2019 Grade 8 B2

Low		Medium		High	
2.4		7.2		12	
1 REP10199-2708				M.8.25:6	
2 REP10199-2861				M.8.15:6	
3 10199-2757				M.8.20:6	
4 REP10199-9677				M.8.13:6	
5 10199-2881				M.8.14:6	
6 10199-2966				M.8.2:6	
7 REP10199-2721				M.8.11:6	
8 10199-2976				M.8.7:6	
9 10199-4332				M.8.17:6	
10 10199-2678				M.8.26:6	
11 10199-3837				M.8.24:6	
12 10199-2880				M.8.21:2	
13 10199-9635				M.8.5:6	
14 REP10199-2768				M.8.14:6	
15 REP10199-2521				M.8.28:6	
16 10199-3839				M.8.4:1	
17 REP10199-2774				M.8.4:6	
18 REP10199-13428				M.8.8:6	
19 10199-2903				M.8.15:6	
20 10199-2919				M.8.10:12	
21 REP10199-3799				M.8.17:6	
22 REP10199-2781				M.8.14:6	
23 10199-2865				M.8.1:6	
24 10199-4239				M.8.22:6	
25 10199-2927				M.8.18:6	
26 10199-2696				M.8.3:6	
27 REP10199-2743				M.8.11:6	
28 REP10199-2857				M.8.5:4	
29 10199-11441				M.8.22:6	
30 10199-2841				M.8.25:12	
31 REP10199-2974				M.8.7:6	
32 10199-2838				M.8.9:6	
33 REP10199-3004				M.8.16:6	
34 10199-2670				M.8.13:6	

Table 8.9
Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])
WV MATH 2019 Grade 8 B2

Low DOK		Matched DOK		High DOK

NS.EE.M.8			
NS.EE.M.8.A			
M.8.1: [1]	23:(6)[1]		
M.8.2: [2]	6:(6)[1]		
NS.EE.M.8.B			
M.8.3: [1]	26:(6)[1]		
M.8.4: [1]	16:(1)[1]	17:(6)[1]	
M.8.5: [2]	13:(6)[1]	28:(4)[1]	
M.8.6: [1]	28:(2)[1]		
NS.EE.M.8.C			
M.8.7: [2]	8:(6)[2]	31:(6)[2]	
M.8.8: [2]	18:(6)[3]		
NS.EE.M.8.D			
M.8.9: [2]	32:(6)[2]		
M.8.10: [2]	20:(12)[2]		
F.M.8			
F.M.8.A			
M.8.11: [1]	7:(6)[1]	27:(6)[1]	
M.8.12			
M.8.13: [2]	4:(6)[1]	34:(6)[1]	
F.M.8.B			
M.8.14: [2]	5:(6)[2]	14:(6)[2]	22:(6)[2]
M.8.15: [2]	2:(6)[2]	19:(6)[2]	
G.SP.M.8			
G.SP.M.8.A			
M.8.16: [2]	33:(6)[2]		
M.8.17: [2]	9:(6)[2]	21:(6)[2]	
M.8.18: [2]	25:(6)[2]		
M.8.19: [2]	16:(5)[1]		

M.8.20: [2]	3:(6)[2]		
G.SP.M.8.B			
M.8.21: [2]	12:(2)[2]		
M.8.22: [2]	12:(4)[1]	24:(6)[1]	29:(6)[1]
M.8.23			
G.SP.M.8.C			
M.8.24: [2]	11:(6)[1]		
G.SP.M.8.D			
M.8.25: [2]	1:(6)[2]	30:(12)[2]	
M.8.26: [2]	10:(6)[2]		
M.8.27			
M.8.28: [2]	15:(6)[2]		
MHM			
MHM1			
MHM2			
MHM3			
MHM4			
MHM5			
MHM6			
MHM7			
MHM8			

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 8B2 MATH

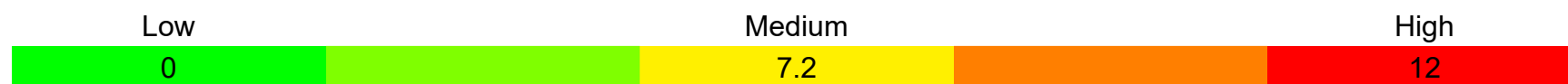
[illegible]

Table 8.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 8B2 MATH

Low		Medium		High	
2.4		7.2		12	
1 REP10199-2708	Exact:5	Partial:1			
2 REP10199-2861	Exact:6				
3 10199-2757	Exact:5	Partial:1			
4 REP10199-9677	Exact:5	Partial:1			
5 10199-2881	Exact:5	Partial:1			
6 10199-2966	Exact:6				
7 REP10199-2721	Exact:6				
8 10199-2976	Exact:5	Partial:1			
9 10199-4332	Exact:6				
10 10199-2678	Exact:6				
11 10199-3837	Exact:5	Partial:1			
12 10199-2880	Exact:3	Partial:3			
13 10199-9635	Exact:6				
14 REP10199-2768	Exact:1	Partial:5			
15 REP10199-2521	Exact:5	Partial:1			
16 10199-3839	Exact:6				
17 REP10199-2774	Exact:6				
18 REP10199-13428	Exact:5	Partial:1			
19 10199-2903	Exact:6				
20 10199-2919	Exact:12				
21 REP10199-3799	Exact:6				
22 REP10199-2781	Exact:5	Partial:1			
23 10199-2865	Exact:6				
24 10199-4239	Exact:6				
25 10199-2927	Exact:6				
26 10199-2696	Exact:6				
27 REP10199-2743	Exact:6				
28 REP10199-2857	Exact:1	Partial:2	Minimal:2	Negligible:1	
29 10199-11441	Exact:6				
30 10199-2841	Exact:12				
31 REP10199-2974	Exact:6				
32 10199-2838	Exact:5	Partial:1			
33 REP10199-3004	Exact:6				
34 10199-2670	Exact:5	Partial:1			

Batch 4 West Virginia Math Grade 8

Table 8.1 *Categorical Concurrence between Standards and Assessment as Rated by Three Reviewers WV MATH 2019 Grade 8 B4 Number of Assessment Items - 34*

Reporting Category			Level by Standards			Hits		Categorical Concurrence
Title	Domain Number	Standard Number	Level	Num of Stds by Level	% w/in RC by Level	Mean	S.D.	
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	4	10	1 2	4 6	40 60	9.25	4.86	YES
F.M.8 Functions (F)	2	5	1 2	1 4	20 80	7.75	2.5	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	4	13	1 2	1 12	7.69 92.31	10.75	5.19	YES
MHM Mathematical Habits of Mind	8	8	2 3	2 6	25 75	0	0	NO
Total	18	36	1 2 3	6 24 6	17 67 17	27.75	12.5	

Table 8.2 *Depth-of-Knowledge Consistency Between Standards as Assessment as Rated by Three Reviewers WV MATH 2019 Grade 8 B4 Number of Assessment Items - 34*

Reporting Category			Hits		DOK Level of Item						DOK Consistency
Title	Domain Num	Std Num	M	SD	% Under	SD	% At	SD	% Above	SD	
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	4	10	9.25	4.86	31.63	17	53.41	10	14.96	11	YES
F.M.8 Functions (F)	2	5	7.75	2.5	34.03	8	54.86	14	11.11	9	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	4	13	10.75	5.19	17.17	17	73.63	22	9.2	11	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	0	0	NT
Total	18	36	27.75	12.5	27.03	8.8	58.56	13.9	14.41	8.3	

Table 8.3 Range-of-Knowledge Correspondence and Balance of Representation between Standards and Assessment as Rated by Three Reviewers WV MATH 2019 Grade 8 B4 Number of Assessment Items - 34

Reporting Category			Hits		Range of Standards				Range of Know	% of Hits of Total Hits		Balance Index		Bal of Rep
					Num Stds Hit		% of Total							
Title	Dom Num	Stds Num	M	S.D	M	S.D	M	S.D		M	S.D	M	S.D	
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	4	10	9.25	4.86	6.25	2.87	62.5	28.72	YES	31	6	0.88	0.08	YES
F.M.8 Functions (F)	2	5	7.75	2.5	4.5	0.58	90	11.55	YES	31	9	0.89	0.09	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	4	13	10.75	5.19	7.75	3.2	59.62	24.63	YES	38	3	0.87	0.09	YES
MHM Mathematical Habits of Mind	8	8	0	0	0	0	0	0	NT	0	0	N/A	0	NT
Total	18	36	27.75	12.5	4.6	3.36	53.03	38		25	17	0.88	0.06	

Table 8.4 Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Three Reviewers WV MATH 2019 Grade 8 B4 Number of Assessment Items - 34

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
NS.EE.M.8 The Number System (NS) and Expressions and Equations (EE)	YES	YES	YES	YES
F.M.8 Functions (F)	YES	YES	YES	YES
G.SP.M.8 Geometry (G) and Statistics and Probability (SP)	YES	YES	YES	YES
MHM Mathematical Habits of Mind	NO	NT	NT	NT

Table 8.5 *Depth-of-Knowledge Levels by Item and Reviewers Intraclass Correlation*
WV MATH 2019 Grade 8 B4

Item	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Reviewer 6
1		2	2	2		1
2		1	1	1		1
3		2	2	2		2
4		1	1	2		2
5		2	2	2		2
6		2	2	2		2
7		1	1	1		1
8		2	2	2		2
9		2	1	2		1
10			1	2		2
11			1	1		1
12			2	2		2
13			2	2		1
14			2	2		1
15			1	1		1
16			2	2		2
17			2	2		1
18			2	2		2
19			1	1		1
20			1	1		1
21			2	2		2
22			2	2		1
23			2	2		1
24			2	1		1
25			2	1		2
26			1	2		2
27			1	2		2
28			2	2		2
29			1	2		2
30			2	2		2
31			3	2		2
32			2	3		3
33			1	1		1
34			2	2		2

Intraclass correlation - .6312

Pairwise Comparison - 0.73

Table 8.6 DOK Levels and Objectives Code by Each Reviewer WV MATH 2019 Grade 8 B4 Number of Reviewers: Three

Number of Reviewers: Six

Item	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj	DOK	Obj	S1 Obj	S2 Obj
1	2	M.8.17			2	M.8.17			2	M.8.17			1	M.8.17		
2	1	M.8.14			1	M.8.14			1	M.8.14			1	M.8.14		
3	2	M.8.28			2	M.8.28			2	M.8.28			2	M.8.28		
4	1	M.8.11			1	M.8.11			2	M.8.11			2	M.8.11		
5	2	M.8.15			2	M.8.15			2	M.8.15			2	M.8.15		
6	2	M.8.12			2	M.8.14			2	M.8.12			2	M.8.12		
7	1	M.8.2			1	M.8.2			1	M.8.2			1	M.8.2		
8	2	M.8.7			2	M.8.7			2	M.8.7			2	M.8.7		
9	2	M.8.21			1	M.8.21			2	M.8.22			1	M.8.22		
10					1	M.8.24			2	M.8.24			2	M.8.24		
11					1	M.8.4			1	M.8.4			1	M.8.4		
12					2	M.8.20			2	M.8.20			2	M.8.20		
13					2	M.8.8			2	M.8.8			1	M.8.8		
14					2	M.8.27			2	M.8.27			1	M.8.27		
15					1	M.8.26			1	M.8.26			1	M.8.26		
16					2	M.8.15			2	M.8.15			2	M.8.15		
17					2	M.8.5			2	M.8.5			1	M.8.5		
18					2	M.8.20			2	M.8.20			2	M.8.20		
19					1	M.8.13			1	M.8.13			1	M.8.13		
20					1	M.8.2			1	M.8.2			1	M.8.2		
21					2	M.8.22			2	M.8.22			2	M.8.22		
22					2	M.8.7			2	M.8.7			1	M.8.7		
23					2	M.8.18			2	M.8.18			1	M.8.18		
24					2	M.8.6			1	M.8.6			1	M.8.6		
25					2	M.8.11			1	M.8.11			2	M.8.11		
26					1	M.8.23			2	M.8.23			2	M.8.23		
27					1	M.8.13			2	M.8.23			2	M.8.13		
28					2	M.8.6			2	M.8.6			2	M.8.6		
29					1	M.8.23			2	M.8.23			2	M.8.23		
30					2	M.8.18			2	M.8.18			2	M.8.18		
31					3	M.8.8			2	M.8.8			2	M.8.8		
32					2	M.8.9			3	M.8.9			3	M.8.9		
33					1	M.8.14			1	M.8.14			1	M.8.13		
34					2	M.8.10			2	M.8.14			2	M.8.10		
Objective Pairwise Comparison: 0.91																
Standard Pairwise Comparison: 0.96																

Table 8.7 Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
WV MATH 2019 Grade 8 B4

	Low 0	Medium 2.4	High 4
NS.EE.M.8			
NS.EE.M.8.A			
M.8.1			
M.8.2	7(4)	20(3)	
NS.EE.M.8.B			
M.8.3			
M.8.4	11(3)		
M.8.5	17(3)		
M.8.6	24(3)	28(3)	
NS.EE.M.8.C			
M.8.7	22(3)	8(4)	
M.8.8	13(3)	31(3)	
NS.EE.M.8.D			
M.8.9	32(3)		
M.8.10	34(2)		
F.M.8			
F.M.8.A			
M.8.11	25(3)	4(4)	
M.8.12	6(3)		
M.8.13	33(1)	19(3)	27(2)
F.M.8.B			
M.8.14	33(2)	34(1)	6(1) 2(4)
M.8.15	5(4)	16(3)	

G.SP.M.8				
G.SP.M.8.A				
M.8.16				
M.8.17	1(4)			
M.8.18	23(3)	30(3)		
M.8.19				
M.8.20	18(3)	12(3)		
G.SP.M.8.B				
M.8.21	9(2)			
M.8.22	9(2)	21(3)		
M.8.23	26(3)	27(1)	29(3)	
G.SP.M.8.C				
M.8.24	10(3)			
G.SP.M.8.D				
M.8.25				
M.8.26	15(3)			
M.8.27	14(3)			
M.8.28	3(4)			
MHM				
MHM1				
MHM2				
MHM3				
MHM4				
MHM5				
MHM6				
MHM7				
MHM8				

Table 8.8

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)
 WV MATH 2019 Grade 8 B4

Low		Medium		High	
0.8		2.4		4	
1 REP10199-3799		M.8.17:4			
2 10199-2725		M.8.14:4			
3 10199-2766		M.8.28:4			
4 10199-2753		M.8.11:4			
5 10199-2911		M.8.15:4			
6 10199-2748		M.8.12:3		M.8.14:1	
7 10199-2970		M.8.2:4			
8 10199-2693		M.8.7:4			
9 10199-4100		M.8.21:2		M.8.22:2	
10 10199-3795		M.8.24:3			
11 10199-2704		M.8.4:3			
12 10199-2810		M.8.20:3			
13 10199-2801		M.8.8:3			
14 10199-2560		M.8.27:3			
15 10199-2937		M.8.26:3			
16 10199-2832		M.8.15:3			
17 10199-2845		M.8.5:3			
18 10199-2811		M.8.20:3			
19 10199-2668		M.8.13:3			
20 10199-2948		M.8.2:3			
21 10199-3739		M.8.22:3			
22 10199-2682		M.8.7:3			
23 REP10199-2920		M.8.18:3			
24 10199-2635		M.8.6:3			
25 10199-2692		M.8.11:3			
26 10199-2891		M.8.23:3			
27 10199-2707		M.8.13:2		M.8.23:1	
28 10199-2963		M.8.6:3			
29 10199-2812		M.8.23:3			
30 10199-9826		M.8.18:3			
31 10199-2712		M.8.8:3			
32 10199-2942		M.8.9:3			
33 10199-2736		M.8.13:1		M.8.14:2	
34 10199-10871		M.8.10:2		M.8.14:1	

Table 8.9

Assessment Item DOK vs Consensus DOK (Item Number: Number of Reviewers [Average DOK])

WV MATH 2019 Grade 8 B4

Low DOK		Matched DOK		High DOK

NS.EE.M.8				
NS.EE.M.8.A				
M.8.1				
M.8.2: [2]	7:(4)[1]	20:(3)[1]		
NS.EE.M.8.B				
M.8.3				
M.8.4: [1]	11:(3)[1]			
M.8.5: [2]	17:(3)[2]			
M.8.6: [1]	24:(3)[1]	28:(3)[2]		
NS.EE.M.8.C				
M.8.7: [2]	8:(4)[2]	22:(3)[2]		
M.8.8: [2]	13:(3)[2]	31:(3)[2]		
NS.EE.M.8.D				
M.8.9: [2]	32:(3)[3]			
M.8.10: [2]	34:(2)[2]			
F.M.8				
F.M.8.A				
M.8.11: [1]	4:(4)[2]	25:(3)[2]		
M.8.12: [2]	6:(3)[2]			
M.8.13: [2]	19:(3)[1]	27:(2)[2]	33:(1)[1]	
F.M.8.B				
M.8.14: [2]	2:(4)[1]	6:(1)[2]	33:(2)[1]	34:(1)[2]
M.8.15: [2]	5:(4)[2]	16:(3)[2]		
G.SP.M.8				
G.SP.M.8.A				
M.8.16				
M.8.17: [2]	1:(4)[2]			
M.8.18: [2]	23:(3)[2]	30:(3)[2]		
M.8.19				

M.8.20: [2]	12:(3)[2]	18:(3)[2]		
G.SP.M.8.B				
M.8.21: [2]	9:(2)[2]			
M.8.22: [2]	9:(2)[2]	21:(3)[2]		
M.8.23: [1]	26:(3)[2]	27:(1)[2]	29:(3)[2]	
G.SP.M.8.C				
M.8.24: [2]	10:(3)[2]			
G.SP.M.8.D				
M.8.25				
M.8.26: [2]	15:(3)[1]			
M.8.27: [2]	14:(3)[2]			
M.8.28: [2]	3:(4)[2]			
MHM				
MHM1				
MHM2				
MHM3				
MHM4				
MHM5				
MHM6				
MHM7				
MHM8				

Number of Reviewers Coding an Item by Objective (Item Number: Number of Reviewers)
CONF WV Grade 8B4 MATH

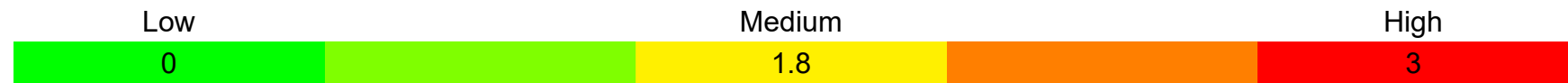
[illegible]

Table 8.CONF

Number of Reviewers Coding an Objective by Item (Objective: Number of Reviewers)

CONF WV Grade 8B4 MATH

Low		Medium		High	
0.6		1.8		3	
1 REP10199-3799	Exact:3				
2 10199-2725	Exact:3				
3 10199-2766	Exact:3				
4 10199-2753	Exact:3				
5 10199-2911	Exact:3				
6 10199-2748	Exact:2			Partial:1	
7 10199-2970	Exact:3				
8 10199-2693	Exact:2			Partial:1	
9 10199-4100	Exact:1			Partial:2	
10 10199-3795	Exact:3				
11 10199-2704	Exact:3				
12 10199-2810	Exact:3				
13 10199-2801	Exact:3				
14 10199-2560	Exact:2			Partial:1	
15 10199-2937	Exact:3				
16 10199-2832	Exact:3				
17 10199-2845	Exact:3				
18 10199-2811	Exact:3				
19 10199-2668	Exact:3				
20 10199-2948	Minimal:1			Negligible:2	
21 10199-3739	Exact:3				
22 10199-2682	Exact:3				
23 REP10199-2920	Exact:3				
24 10199-2635	Exact:3				
25 10199-2692	Exact:3				
26 10199-2891	Exact:3				
27 10199-2707	Exact:3				
28 10199-2963	Exact:3				
29 10199-2812	Exact:3				
30 10199-9826	Exact:3				
31 10199-2712	Exact:3				
32 10199-2942	Exact:3				
33 10199-2736	Exact:2			Partial:1	
34 10199-10871	Exact:3				

**Information subject to nondisclosure agreements has
been redacted for public release.**

Appendix C

Reviewers' Notes

West Virginia College- and Career- Readiness Standards for Mathematics Grades 3-8 and Sample Test Events

2019

**Information subject to nondisclosure agreements has
been redacted for public release.**

Appendix D

Debriefing Summary Notes

West Virginia College- and Career- Readiness Standards for Mathematics Grades 3-8 and Sample Test Events

2019

Appendix E

DOK Definitions for Mathematics

2019

DOK Primer

Dr. Norman Webb originally developed the Depth of Knowledge (DOK) language system for the purpose of evaluating the relationship between the content complexity (also often referred to as “cognitive complexity”) of curriculum standards and of corresponding assessment items. Over the years, the use of DOK has extended far beyond the original context of evaluating the alignment of statewide summative assessments. Now used extensively throughout U.S. school districts, by state departments of education, assessment developers, educational publishers, and others, DOK is applied toward informing alignment between and among all areas of the education system, not just the relationship between standards and assessments. DOK is a tool that allows educators to communicate effectively, consistently, and efficiently about the content complexity of standards, learning objectives, tasks, prompts, questions, etc.

What DOK is	What DOK is NOT
- DOK is an evaluative tool used for content analysis. Specifically, DOK is a language system that can be used to differentiate between and among different levels of complexity of student engagement expressed within components of educational materials/systems. - DOK can be used to interpret standards, learning objectives, tasks, prompts, questions, etc. - DOK is a reflective lens used to foster intentionality in teachers’ practices, to help ensure that the complexity of expected learning outcomes are clearly understood, that (formative/summative/etc.) assessments provide opportunities to make reasonable inferences about attainment of learning outcomes, and that appropriate educational opportunities are provided to allow students to engage at the level(s) of complexity intended. - DOK is a conceptualization of complexity, as differentiated from difficulty.	- DOK is not used to evaluate text complexity, topic complexity (e.g. how complex is photosynthesis), phenomenon complexity (NGSS), or overall complexity of, for example, an entire lesson or unit. - DOK is not a rubric. - DOK is not a protocol. - DOK is not a type of question (The idea of “DOK questions” is not consistent with the intent of DOK.) - DOK is not a measurement of “how deeply” an individual is engaging with a topic. - DOK is not hierarchical or progressive (i.e. it does not reflect any sort of learning progression from low→ high complexity). - DOK is not a value judgment and does not reflect importance. (In other words, there is no idea inherent to DOK that any level of DOK is “better” than any other. The standards, by definition, specify what is important.)

Selected Examples of how DOK Is Commonly Used

- By individual K-12 teachers and within school Professional Learning Communities (PLCs) to understand the meaning of academic standards – e.g. the complexity at which students are expected to engage with a concept or idea.
- By K-12 teachers and associated C&I staff to inform development of lesson plans, unit designs, formative/summative classroom and district assessments, and other materials.
- By K-12 teachers and associated C&I staff to inform selection of assessment items from item bank products for which a district has purchased access.
- To communicate expectations to item writers / content developers – e.g. expected distribution of DOK levels of items on an assessment.
- In specifications for assessments via state RFPs.
- By large-scale state department of education efforts to provide educators with tools for goals of school improvement, assessment development, and other endeavors.
- As one component of alignment analyses of statewide summative assessments with standards (as specified by “depth and breadth” requirement of ESSA).

Point of clarification: We use DOK to do *content analyses* of the “content” of standards, tasks, instructional prompts, questions, etc. The “content” of a standard can be thought of as “the entirety of the standard.” Hence, when we use the term “content complexity” we are referring to the “content” within a learning expectation (standard) or a task, prompt, question, etc. We can analyze the complexity of the “content” of these expectations/tasks/etc. (everything that is within) that we read on paper (or on screen). The term “content complexity” does NOT refer to the complexity of a topic (e.g. photosynthesis), text, or context, etc.



Mathematics DOK Definitions

DOK 1 (Recall)

DOK 1 is defined by the rote recall of information or performance of a simple, routine procedure. For example, repeating a memorized fact, definition, or term, performing a simple algorithm, rounding a number, or applying a formula are DOK 1 performances.

Performing a one-step computation or operation, executing a well-defined multi-step procedure or a direct computational algorithm are also included in this category. Examples of well-defined multi-step procedures include finding the mean or median or performing long division. Reading information directly from a graph, plugging data into an electronic device to derive an answer, or simple paraphrasing are all tasks that are considered a level of complexity comparable to recall. A student answering a Level 1 item either knows the answer or does not: that is, the item does not need to be “figured out” or “solved.”

At a DOK 1, problems in context are straightforward and the solution path is obvious. For example, the problem may contain a keyword that indicates the operation needed. Other DOK 1 examples include plotting points on a coordinate system, using coordinates with the distance formula, or drawing lines of symmetry of geometric figures.

At more advanced levels of mathematics, symbol manipulation and solving a quadratic equation or a system of two linear equations with two unknowns are considered comparable to recall assuming students are expected or likely to use well-known procedures (e.g. factoring, completing the square, substitution, or elimination) to derive a solution. Operating on polynomials or radicals, using the laws of exponents, or simplifying rational expressions are considered rote procedures.

Verbs should not be classified as any category without considering what the verb is acting upon or the verb’s direct object. “*Identify* attributes of a polygon” is recall, but “*identify* the rate of change for an exponential function” requires a more complex analysis. To *describe* by listing the steps used to solve a problem is recall (i.e, *Show your work*) whereas to *describe* by providing a mathematical argument or rationale for a solution is more complex.

Webb, N. L. *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states*. A study of the State Collaborative on Assessment & Student Standards (SCASS) Technical Issues in Large-Scale Assessment (TILSA). Washington, D. C.: Council of Chief State School Officers, December 2002. Revised in 2014 by Norman Webb, Sara Christopherson with the help of Lynn Raith.

WebbAlign offers alignment studies and DOK professional development training. Please contact WCEPS at webbalign@wceps.org or 1-877-249-4211 for more information.

DOK 2 (Skill/Concept)

DOK 2 involves engaging in some mental processing beyond a habitual response as well as decision-making about how to approach the problem or activity. This category can require conceptual understanding and/or demonstrating conceptual knowledge by explaining thinking in terms of concepts.

DOK 2 tasks includes distinguishing among mathematical ideas, processing information about the underlying structure, drawing relationships among ideas, deciding among and performing appropriate skills, applying properties or conventions within a relevant and necessary context, transforming among different representations, interpreting and solving problems and/or graphs. When given a problem statement, formulating an equation or inequality, deriving a solution, and reporting the solution in the context of the problem fit within DOK 2. Processes such as classifying, organizing, and estimating that involve attending to multiple attributes, features, or properties also fall into this category.

Verifying that the number of objects in one set is larger or fewer than the number of objects in a second set by matching pairs or forming equivalent groups is a DOK 2 activity for a kindergartener. A first grader modeling a joining or separating situation pictorially or physically also is in this category.

Skills and concepts include constructing a graph and interpreting the meaning of critical features of a function, beyond just identifying or finding such features as well as describing the effects of parameter changes. Note, however, that using a well-defined procedure to find features of a standard function, such as the slope of a linear function with one variable or a quadratic, is a DOK 1. Graphing higher order or irregular functions is a DOK 2. Basic computation, as well as converting between different units of measurement, are generally a Category 1, but illustrating a computation by different representations (e.g. equations and a base-ten model) to explain the results is a DOK 2. Computing measures of central tendency (applying set procedures) is a DOK 1, but interpreting such measures for a data set within its context or using measures to compare multiple data sets is a DOK 2. Performing original formal proofs is beyond DOK 2, but explaining in one's own words the reasons for an action or application of a property is comparable to a DOK 2. Activities at a DOK 2 are not limited only to number skills, but may involve visualization skills (e.g. mentally rotating a 3D figure or transforming a figure) and probability skills requiring more than simple counting (e.g. determining a sample space or probability of a compound event). Other activities at this category include detecting or describing non-trivial patterns, explaining the purpose and use of experimental procedures, and carrying out experimental procedures.

DOK 3 (Strategic Thinking)

DOK 3 requires reasoning and analyzing using mathematical principles, ideas, structure, and practices. DOK 3 includes solving involved problems; conjecturing; creating novel solutions and forms of representation; devising original proofs, mathematical arguments, and critiques of arguments; constructing mathematical models; and forming robust inferences and predictions. Although DOK 2 also involves some problem solving, DOK 3 includes situations that are non-routine, more demanding, more abstract, and more complex than DOK 2. Such activities are characterized by producing sound and valid mathematical arguments when solving problems, verifying answers, developing a proof, or drawing inferences. Note that the sophistication of a mathematical argument that would be considered DOK 3 depends on the prior knowledge and experiences of the person. For example, primary school student arguments for number problems can be a DOK 3 activity (e.g. counting number of combinations, finding shortest route from home to school, computing with large numbers) as can abstract reasoning in developing a logical argument by students in higher grades. DOK 3 problems are those for which it is not evident from the first reading what is needed to derive a solution and so require demanding reasoning to work through. Such problems usually can be solved in different ways and may even have more than one correct solution based on different stated assumptions. Paraphrasing in one's own words or reproducing a proof that was previously demonstrated is a DOK 2. Applying properties and producing arguments in proving a theorem or identity not previously seen is a DOK 3. Also in the DOK 3 category is making sense of the mathematics in a situation, creating a mathematical model of a situation considering contextual constraints, deriving a new formula, designing and conducting an experiment, and interpreting findings.

DOK 4 (Extended Thinking)

DOK 4 demands are at least as complex as those of DOK 3, but a main factor that distinguishes the two categories is the need to perform activities over days and weeks (DOK 4) rather than in one sitting (DOK 3). The extended time that accompanies this type of activity allows for creation of original work and requires metacognitive awareness that typically increases the complexity of a DOK 4 task overall, in comparison with DOK 3 activities. Category 4 activities require complex reasoning, planning, research, and verification of work. Conducting a research project, performance activity, an experiment, and a design project as well as creating a new theorem and proof fit under Category 4. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, collecting water temperature from a river each day for a month and then reporting the findings by constructing a graph is a DOK 2 activity. Developing a mathematical model of the flow of water in a river for all four seasons using a number of variables would be a DOK 4 activity. It is likely that a DOK 4 activity will require making connections among a number of ideas or variables within the area of mathematics or among a number of content areas. Category 4 activities require selecting an appropriate approach among many alternatives to produce a product, conclusion, or finding, such as critiquing a body of work, synthesizing ideas in a new way, or creating an original model.

General Guidelines for Assigning DOK:

- The DOK definitions can be applied to mathematics standards, tasks, or activities.
- Consider the complexity of the mathematical demands, not the difficulty for students.
- Consider the mathematical experience (prior knowledge) and grade-level expectations of a typical student.
- Do not rely on verbs (describe, explain, evaluate, etc.). Instead, consider the content complexity required for an adequate response.
- For multiple-choice assessment items, consider the item as a whole—including distractors—to judge complexity.
- An expectation or item that is confusing due to error or wording does not reflect increased content complexity—it simply means the statement needs revisions.

Appendix F

Coding Instructions Provided to Panelists

2019

Coding How-To/Reminders for Reviewers

WV Alignment Study, June, 2019

STEP ONE: INDEPENDENT ANALYSIS

- Identify what knowledge is **necessary and sufficient** to answer the item correctly. Code item to corresponding standard, choosing the most specific standard possible.
- ONLY if the **necessary and sufficient** knowledge is expressed in more than one standard, code a primary standard and one or two secondary standard(s).
- The content of an item may not correspond to a particular standard but still fit in that “generic” level. Code to the closest generic standard and add a reason why the item was coded to that generic.
- The standards apply to all students; think about what the “typical” grade-level student will need to do to respond to a prompt.
- Consider the full scope of standards.
- Always refer to the DOK definitions to ensure consistent coding.
- Write a reason for any item that contains a Source-of-Challenge issue. Use only in fairly extreme situations when items are technically flawed – i.e. could be answered correctly for the wrong reason or incorrectly for the wrong reason.
- Other comments about strong items, weak items, or items that are perceived as “good” or “bad” should go into the Notes box.
- Write brief comments on any item, but don’t slow down the coding process.
- Record your standard and DOK assignments on both laptops...just in case you get off-number.
- SAVE, SAVE, SAVE!

Debriefing Questions:

- Responses to the debriefing questions allow study directors to obtain a greater sense of the qualitative topics and issues associated with that particular grade-level analysis.
- Do not repeat any information already entered in DOK/standard coding.
- Add feedback or other information not captured in the DOK/standard data.

STEP TWO: CONFIRMATORY ANALYSIS

For each item, consider the internally coded standard(s).

To what extent does the item assess the content (expectations) within the internally coded standard(s)?

- **EXACTLY** (Note that the item does NOT need to assess every aspect of a standard, but it needs to be a direct (“exact”) measurement of a central aspect of the standard. A correct response to the item allows for a direct inference about student knowledge/skills/abilities as expressed in the standard.)
- **PARTIALLY** (The item somewhat targets the expectations within the standard and it can be considered a majority match. A correct response to the item allows for some inference about student knowledge/skills/abilities as expressed in the standard.)
- **MINIMALLY** (The item only very minimally targets the expectations within the standard – and it can be considered only a minority match. A correct response to the item allows for very little or very indirect inference about student knowledge/skills/abilities as expressed in the standard.)
- **NOT AT ALL** (The item does not assess the expectations within the standard. No inference can be made about student knowledge/skills/abilities as expressed in the standard based on a correct response.)

For each item, consider the internally assigned DOK.

- **If you agree with the internally assigned DOK**, enter that value (1, 2, or 3) into the drop-down menu on the WATv2 for that item.
- **If you do not agree with the internally assigned DOK**, enter the DOK that you assigned to the item.



Device Comparability

This document provides evidence of the comparability of the West Virginia General Summative Assessment (WVGSA) tests across the most frequently used platforms (e.g., computers, tablets).

American Institutes for Research (AIR), the test vendor for the WVGSA, maintains an indicator in the data showing the types of devices used by test takers. This indicator can be used alongside the observed test scores to assess the degree to which WVGSA scores are comparable for students on different devices.

Study Background

Score comparability across different devices can be examined to assess whether student performance on the WVGSA differs among students, given the prior year's test scores. For example, the data can be used to examine if students who take the WVGSA test on a tablet tend to have higher or lower scores than students who take the test on a Chromebook or a Windows PC. If there is a device effect (e.g., systematically lower or higher scores on a certain device, relative to other devices), it may suggest that students taking the test on that device have a disadvantage or advantage that is affecting their scores.

This lends itself to a research question: “Are scores comparable for students participating in the WVGSA from any device used for test administration?” Simply examining the current year's scores and disaggregating by device would be insufficient. Students are not randomly assigned to different devices, so we must control for the potential effects of any preexisting differences that could confound the outcomes.

This study analyzes WVGSA data for students participating in the mathematics and English language arts (ELA) grade 7 administration and evaluates the degree to which scores are comparable for students taking the test across devices. The approach controls for preexisting differences between students to control for the non-random assignment of students to different devices.

Methods

The device comparability study uses grade 7 mathematics scaled scores from spring 2019 and controls for preexisting differences among students using grade 6 mathematics test scores from the same students in the previous year as a covariate; the same process was implemented for ELA. An indicator for device is available at the student level, so we can apply a regression model with the following form

$$y_{tig} = \mu + \beta_1(y_{t-1,i}) + \sum_{j=2}^4 \beta_j(D_j) + \delta_g + \varepsilon_{ig}$$

where y_{tig} is the grade 7 mathematics or ELA scaled score for the i th student who is in school g at time t . The coefficient β_1 is the effect of the prior year grade 6 mathematics or ELA score and is used to control for preexisting differences between individuals. This variable is measured with error, so the model accounts for that error to avoid bias as described by Doran (2014) and Greene (2000). The coefficients β_j for $j = \{2, \dots, 4\}$ represent the effect of device D , which is a binary coded variable indicating that

$$D_j = \begin{cases} 1 & \text{if student was administered device } j \\ 0 & \text{otherwise} \end{cases}$$

Students are clustered into common groups, so the random effect δ_g is used to account for clustering at the school level to return model-based standard error consistent with the clustered nature of the data.

We can examine the device effects marginally via the regression coefficients. However, that doesn't answer the overall research question. To broadly determine if use of a specific device leads to scores that are significantly higher or lower than any other devices, we can use a likelihood ratio test (LRT) to compare the model expressed above (now referred to as the fully specified model) to a baseline (reduced) model that has the simple form

$$y_{tig} = \mu + \beta_1(y_{t-1,i}) + \delta_g + \varepsilon_{ig}.$$

This equation represents a reduced form of the model above in which the only difference is that the set of device predictors is not included. The deviance between the fully specified model and the reduced model can be compared using the LRT as the overall omnibus test to assess whether any device is significantly different from any other device overall.

The LRT test used is

$$Deviance = 2[LL_{fs} - LL_r].$$

The deviance is a χ^2 distributed variable with degrees of freedom equal to the difference in the number of parameters. The p -value of the deviance serves as an indicator of the degree to which the fully specified model is significantly different from the reduced model. If the p -value of the difference between the two models is significant, then it suggests that student scores on at least one of the devices are significantly different from scores on one of the other devices. If the p -value on the deviance is not significant, it indicates that scores between devices are comparable.

Results

The data used for this study are summarized in Table 1 and Table 2, which show the numbers of students used in this analysis. Table 3 and Table 4 provide the numbers (N) of students by device and the means and standard deviations of the scores disaggregated by device for the common sample of students between 2019 and 2018. Table 5 and Table 6 provide the results of the regression models showing the model coefficients and their standard errors for the fully specified and reduced models. Lastly, Table 7 and Table 8 provide the results of the LRT, showing the difference between the effects of the two models. The LRT was not performed for ELA because adding the device variable into the baseline model did not reduce the log-likelihood.

Table 1: Sample Size by Administration, Mathematics

Administration	N
Spring 2019 Grade 7	18,853
Spring 2018 Grade 6	19,125
Common between administrations	17,900

Table 2: Sample Size by Administration, ELA

Administration	N
Spring 2019 Grade 7	18,840
Spring 2018 Grade 6	19,085
Common between administrations	17,886

Table 3: Sample Size by Device, Grade 7 Mathematics, Spring 2019

Device	N	Scale Score	
		Mean	SD
Windows	10,815	522.32	63.56
Chrome	4,052	525.44	63.97
Mac	456	515.73	64.12
iPad	2,577	529.35	67.76
Total	17,900	523.87	64.34

Table 4: Sample Size by Device, Grade 7 ELA, Spring 2019

Device	N	Scale Score	
		Mean	SD
Windows	10,767	630.07	47.78
Chrome	4,083	634.13	48.58
Mac	387	627.00	51.47
iPad	2,649	633.15	49.49
Total	17,886	631.39	48.34

Table 5: Regression Coefficients, Mathematics

Model	Variable	Coefficient	Standard Error
Baseline Model	(Intercept)	-45.096	3.704
	Prior Year Grade 6 Scale Score	1.148	0.007
Fully Specified Model	(Intercept)	-45.363	3.974
	Prior Year Grade 6 Scale Score	1.147	0.007
	iPad	4.749	2.578
	Mac	-1.436	3.423
	Windows	-0.0427	1.552

Table 6: Regression Coefficients, ELA

Model	Variable	Coefficient	Standard Error
Baseline Model	(Intercept)	44.037	3.662
	Prior Year Grade 6 Scale Score	0.934	0.005
Fully Specified Model	(Intercept)	43.439	3.813
	Prior Year Grade 6 Scale Score	0.934	0.005
	iPad	0.782	1.628
	Mac	-0.868	2.347
	Windows	0.680	1.065

Table 7: Log Likelihood and Deviance Statistics, Mathematics

Model	Log Likelihood	Deviance	DF	p-value
Baseline Model	-89617.65	5.44	3	0.142
Fully Specified Model	-89614.93			

Table 8: Log Likelihood and Deviance Statistics, ELA

Model	Log Likelihood	Deviance	DF	p-value
Baseline Model	-84860.10	-2.94	3	-
Fully Specified Model	-84861.57			

Summary

The p -value for the Math LRT test is non-significant, indicating that the fully specified model adds no predictors that are significantly different from the reduced model. For ELA, adding device variables into the baseline model did not improve the model fit, suggesting that there is no difference between fully specified and reduced model. These results indicate that all regression coefficients on the devices are statistically equivalent, meaning that there are no statistically significant differences among the scores for students participating in the WVGSA on the different types of devices. The data support the notion that no systematic differences exist in the scores for students when administered the WVGSA on different devices.

References

Doran, H.C. (2014), Methods for Incorporating Measurement Error in Value-Added Models and Teacher Classifications, *Statistics and Public Policy* 1(1), 114–119.

Greene, W.H. (2000), *Econometric Analysis (4th ed.)*, Saddle River, NJ: Prentice Hall.

Special Study: Differential Item Functioning Across Students with Accommodations for the Independent College and Career Readiness Item Bank

Within the last year, the West Virginia General Summative Assessment (WVGSA) underwent peer review by the US Department of Education. Following this evaluation, evidence for the fairness and validity of the use of test accommodations was requested. To provide additional evidence regarding equitable score interpretation and comparisons between students utilizing accommodations and those who do not require them, the current study used differential item functioning (DIF) analyses to evaluate the presence of any systematic bias against groups of students who utilize accommodations.

The WVGSA is one of the state-wide assessments that use the Cambium Assessment's Independent Career and College Readiness (ICCR) item bank. During spring summative assessments, newly developed items are concurrently field-tested across the states that share the ICCR item bank, namely North Dakota, New Hampshire, West Virginia, and Wyoming. To overcome the limitations in a DIF analysis that would be caused by insufficient number of students in the group using accommodations when only WVGSA data is used, the current study used students' responses (scores) to those common items obtained across these four states.

The first section of this report includes the types of accommodations available for students across the four states. The second section summarizes the methods and procedures used in this study. The results of the analysis and discussions of the results are presented in the third section followed by the conclusions in the fourth section.

Types of Accommodations Available for Students

Table 1 (adopted from WVGSA Technical Report, Vol 5, p. 4) presents three categories of accommodation features that are available for States to include in their individual assessment programs. Universal tools are available for all students. Students with a plan that is established by an IELP or intervention team may use the feature(s) in the category of designated supports. Students with an IEP or Section 504 plan may use the feature(s) in the category of accommodations. Not all states opt to include the same features. These supports may be digitally delivered (embedded) through the Test Delivery System (TDS) or supplied as a separate supplement to the student (nonembedded). Volumes 1 and 5 of the WVGSA technical manual include details on these features and how test administrators are trained in their use.

Table 1. Universal Tools, Designated Supports, and Accommodations Available in Spring 2019

Universal Tools		Accommodations
Embedded	Breaks	
	Desmos Calculator	
	Dictionary	American Sign Language
	Expandable Items	Audio Translations
	Expandable Passages	Braille
	Highlighter	Closed Captioning
	Keyboard Navigation	Color Choices
	Line Reader	Emboss
	Mark for Review	Mouse Pointer
	Notepad–Digital	Permissive Mode
	Notes–Global	Streamline
	Spell Check	Text-to-Speech
	Strikethrough	Translations-Stacked
	Thesaurus	
	Zoom	
	Non-embedded	
		American Sign Language
		Braille
		Calculator
Dictionary		Color Overlay
Scratch and/or Graph Paper		Magnification
Thesaurus		Multiplication Table
		Noise Buffers
		Print-on-Demand
		Read-Aloud-English
		Scribe
		Speech-to-Text
Paper-Pencil		100s Number Table
		American Sign Language
		Braille
		Calculator
	Breaks	Color Overlay
	Calculator	Dictionary
	Dictionary	Glossary Translations
	Glossary	Large Print
	Line Reader	Magnification
	Scratch and/or Graph Paper	Masking
	Thesaurus	Multiplication Table
		Noise Buffers
		Online Fixed-Form Math
	Read-Aloud	
	Scribe	

Universal Tools	Accommodations
	Speech-to-Text Translated— Test Directions

Method

Differential Item Functioning. Differential Item Functioning (DIF) occurs when students with the same ability level but from two different groups have a different probability of providing a correct response to an item. In other words, members of one group do not have an equal chance of getting a correct score although their ability levels are comparable to the other group. More details of DIF analysis can be found in WVGSA Technical Report, Vol 1. Identifying DIF in the current study is important because it provides a statistical flag for items that may function differently if students use accommodation feature(s) to solve the items.

The current study applied a generalized Mantel-Haenszel (MH) procedure to detect DIF (see Technical Manual, Volume 1 for complete details). The focal group was comprised of those students who used the accommodation under investigation (accommodation group), and the reference group were those students who did not require them (non-accommodation group). DIF-flagged items are classified into three categories of DIF (A, B, or C), with A indicating there was no presence or a negligible amount of DIF, B a moderate amount, and C a moderate to large amount (Zwick, 2012). Items are further identified as having positive DIF, favoring the focal group, or negative DIF, favoring the reference group. If an item falls into category C for any group, the item is reviewed for potential content bias. Table 2 presents the classification rules for dichotomous and polytomous items separately.

Table 2: DIF Classification Rules

Dichotomous Items	
Category	Rule
C	MH_{χ^2} is significant, and $ \hat{\Delta}_{MH} \geq 1.5$.
B	MH_{χ^2} is significant, and $1 \leq \hat{\Delta}_{MH} < 1.5$.
A	MH_{χ^2} is not significant, or $ \hat{\Delta}_{MH} < 1$.
Polytomous Items	
Category	Rule
C	MH_{χ^2} is significant, and $ SMD /SD > .25$.
B	MH_{χ^2} is significant, and $.17 < SMD /SD \leq .25$.
A	MH_{χ^2} is not significant, or $ SMD /SD \leq .17$.

Data. The samples for the current study were drawn from the four spring 2019 summative assessments: the WVGSA, the North Dakota Summative Assessment (NDSA), the Wyoming Test of Proficiency and Progress (WY-TOPP), and the New Hampshire Statewide Assessment

System (NHSAS). Accommodation types available for students and the number of test sessions in which the type of accommodation was utilized are summarized in Tables 3 and 4 for ELA and mathematics, respectively. Note that Tables 2 and 3 list the number of test sessions in which each accommodation was used individually across all persons and all items across four states. These numbers do not reflect the number of responses to individual items within each group (i.e., accommodation vs. non-accommodation). Appendix A presents the number of test sessions for each individual State as well as their individual lists of available accommodations.

Table 3. Number of Test Sessions with Embedded and Non-Embedded Accommodations Across ND, NH, WV, and WY - ELA

Accommodations	Grade					
	3	4	5	6	7	8
Embedded						
American Sign Language	7	13	11	12	6	8
Braille	6	3	9	3	4	7
Streamlined Mode	77	58	58	50	58	76
Text-to-Speech: Items	1,734	2,034	2,206	2,206	2,161	1,998
Text-to-Speech: Passages	3	6	9	1	1	1
Text-to-Speech: Passages & Items	3,598	4,030	4,230	3,910	3,761	3,480
TTS Tracking (WY Only)	405	416	449	321	396	313
TTS (NH Only)	1549	1516	1358	1222	986	883
Non-Embedded						
Alternate Response Options	-	2	4	-	2	1
Print-on-Demand: Stimuli and Items	41	50	28	31	28	24
Read-Aloud Stimuli	708	762	700	613	467	423
Speech-to-Text	214	241	263	223	205	172

Note: Grade 9 and 10 are excluded in this summary as the total number of students across States was insufficient for analysis.

Table 4. Number of Test Sessions with Embedded and Non-Embedded Accommodations Across ND, NH, WV, and WY - Mathematics

Accommodations						
	3	4	5	6	7	8
Embedded Accommodations						
American Sign Language	1	5	1	4	1	2
Braille	4	3	2	3	3	6
Streamlined Mode	99	101	79	78	79	100

Color Choices	1	3	2	7	0	1
Mouse Pointer	1	0	0	1	0	0
Zoom	27	37	15	19	7	15
Text-to-Speech: Items	697	852	767	722	674	596
Text-to-Speech: Passages	1	-	-	-	-	-
Text-to-Speech: Passages & Items	711	-	-	-	-	-
Text-to-Speech: Stimuli	-	1	-	1	-	-
Text-to-Speech: Stimuli & Items	-	758	729	733	686	651
TTS Tracking	459	513	466	367	411	381
Text-to-Speech Sum	1,409	1,611	1,496	1,456	1,360	1,247
Text-to-Speech (NH Total)	1,654	1,831	1,557	1,558	1,268	1,191
Text-to-Speech: Instructions, Passages & Items (WV Total)	2456	2791	2943	2739	2540	2463
Non-Embedded Accommodations						
Alternate Response Options	-	1	4	1	3	1
Print-on-Demand: Stimuli and Items	74	75	46	50	68	45
Read-Aloud Stimuli	466	490	379	265	186	155
Speech-to-Text	177	250	243	208	188	174
Calculator	0	0	0	7094	7720	7677
100s Number Table	0	413	362	323	237	162
Multiplication Table	-	276	385	487	486	432
Abacus	9	4	4	4	2	12
ASL Human Signer	3	4	-	-	-	-

Note: Grade 9 and 10 are excluded in this summary as the total number of students across States is insufficient for analysis.

Item Selection for DIF Analysis. A minimum sample size of 200 responses (Zwick, 2012) per item in each group of students is required for DIF analyses, therefore only items with equal to or greater than 200 student responses were selected for inclusion in the study. Among the accommodation types identified in Table 2 and 3, the subtype of Text-to-Speech in ELA and calculator in grade 6 mathematics were the only types that yielded items with greater than 200 responses. Consequently, a DIF analysis was conducted for those subtypes for the selected grades. Note that the non-embedded accommodation in Appendix B includes all subtypes of non-embedded accommodation types listed in Table 3 and 4.

DIF Analysis Results

Table 5 presents the key results of DIF analyses summarizing the number of items flagged rated with a C level DIF. Appendix B presents the percentage of items rated with other DIF categories. For both ELA and mathematics, the percent of items classified with a C rating was found to be low. For mathematics, 14 items were found to exhibit significant DIF, or a C- rating, with the percentage of items ranging from 1.15% to 6.6% across grades. For ELA, only one item in the 7th grade Text-to-Speech category was found to exhibit significant DIF, with the reference group performing slightly better than the focal group.

Table 5 Number of Items Rated with C Level DIF

Grade	Accommodation Type	Average number of Students per item	Number of Items	Number of Items with C+	Number of Items with C – (%)
ELA					
7	Text-to-Speech	425	73	0	1 (1.37%)
Math					
3	Text-to-Speech	368	93	0	5 (5.37%)
4	Text-to-Speech	374	106	0	7 (6.6%)
5	Text-to-Speech	443	80	0	1 (1.25%)
8	Text-to-Speech	312	87	0	1 (1.15%)

Item Review. It is important to note that when items are found to contain significant DIF, this does not necessarily provide evidence for test bias. Additional review by content experts is necessary to determine the likely cause of the differences in item behavior between the groups. 15 items flagged with a C rating were sent for review by the Cambium Assessment content team. In their evaluation, the content experts were unable to identify any reason that these items might be functioning differently due to the accommodations in question and made the determination that the items should be retained.

Discussion

DIF analysis is an important piece of evidence regarding the fairness and accessibility of any assessment, and ideally, the number of items found to exhibit DIF will always be negligible. In the current study, one ELA item and 14 mathematics items were identified as significantly favoring the reference group. However, the extremely low percentage of items found to contain DIF across these groups indicate that overall, students who utilize these accommodations have the same probability of a correct response as those students who have no need of accommodations on the majority of items in the item pool. Additionally, the content experts were unable to identify any reason that the accommodations themselves might be responsible for this difference in the items' behavior. With this, it is concluded that the scores obtained by students who utilize accommodations while taking the WVGSA are equitable to the scores obtained by students without the use of accommodations.

References

- North Dakota Department of Public Instruction (2019). *North Dakota State Assessment for English Language Arts/Literacy and Mathematics, 2018 – 2019, Volume 1, Annual Technical Report*.
- New Hampshire Department of Education (2019). *New Hampshire Statewide Assessment System, 2018 – 2019, Volume 1, Annual Technical Report*.
- West Virginia Department of Education (2019). *West Virginia General Summative Assessment, 2018 – 2019, Volume 1, Annual Technical Report*.
- Wyoming Department of Education (2019). *Wyoming Test of Proficiency and Progress (WVGSA), 2018 – 2019, Volume 1, Annual Technical Report*.
- Zwick, R. (2012). A Review of ETS Differential Item Functioning Assessment Procedures: Flagging Rules, Minimum Sample Size Requirements, and Criterion Refinement (ETS Research Report No. 12-08). Princeton, NJ: Educational Testing Service.

Appendix A.

The number of test sessions with embedded and non-embedded accommodations by State - ELA

North Dakota (adopted from NDSA Technical Report, Vol 1, p. 12)

Accommodations	Grade					
	3	4	5	6	7	8
Embedded Accommodations						
American Sign Language	-	5	1	3	-	2
Braille	2	-	-	-	1	1
Streamlined Mode	12	9	10	4	31	20
Text-to-Speech: Items	108	88	79	85	57	26
Text-to-Speech: Passages	1	3	1	1	1	-
Text-to-Speech: Passages & Items	649	740	691	768	724	670
Non-Embedded Accommodations						
Alternate Response Options	-	2	4	-	2	1

Print-on-Demand: Stimuli and Items	5	10	3	10	10	7
Read-Aloud Stimuli	341	418	378	371	282	289
Scribe Items (Writing)	165	232	219	214	136	128
Speech-to-Text	69	97	103	122	103	106

New Hampshire (adopted from NHSAS Technical Report, Vol 1, p. 8)

Accommodations	Grade					
	3	4	5	6	7	8
Embedded Accommodations						
American Sign Language	3	3	2	-	-	2
Braille	-	-	-	-	-	2
Emboss: Stimuli & Items	-	-	-	-	-	2
Streamlined Mode	65	49	48	46	27	56
Text-to-Speech	1,549	1,516	1,358	1,222	986	883
Non-Embedded Accommodations						
Abacus	-	-	-	-	-	3
ASL Human Signer	-	1	-	-	-	-
Print-on-Request	32	40	24	20	17	16
Read Aloud	367	344	322	242	185	134
Scribe	319	336	228	191	153	113
Speech-to-Text	145	144	160	101	102	66

West Virginia (adopted from WVGSA Technical Report, Vol 1, p. 10)

Accommodations	G3	G4	G5	G6	G7	G8
Embedded Accommodations						
American Sign Language	4	5	8	8	6	4
Audio Transcriptions	1	6	9	13	6	15
Braille	3	3	1	2	2	2
Color Choices	3	3	2	9	0	17
Mouse Pointer	1	0	0	2	0	1

Text-to-Speech: Instructions, Passages & Items	2,457	2,797	2,946	2,741	2,539	2,464
Text-to-Speech: Instructions & Items	1,484	1,796	1,981	1,907	1,812	1,751
Non-Embedded Accommodations						
Print-on-Demand: Stimuli & Items	4	-	1	1	1	1

Wyoming (adopted from WY-TOPP Technical Report, Vol 1, p. 9)

Accommodations	Grade					
	3	4	5	6	7	8
Embedded Accommodations						
American Sign Language	-	-	-	1	-	-
Braille	1	-	-	1	-	1
Closed Captioning	8	4	8	10	10	11
Print-on-demand	35	28	29	13	39	22
Streamlined Mode	14	5	2	2	4	3
TTS: Items	142	150	146	169	129	143
TTS: Passages	2	3	8	-	-	1
TTS: Passages & Items	492	493	593	401	498	346
TTS Tracking	405	416	449	321	396	313
Zoom	29	37	16	20	8	15

**Note: No data available for Non-Embedded Accommodations for WY*

**The number of test sessions with embedded and non-embedded accommodations by State –
Mathematics**

North Dakota (adopted from NDSA Technical Report, Vol 1, p. 13 - 14)

Accommodations	Grade					
	3	4	5	6	7	8
Embedded Accommodations						
American Sign Language	-	5	1	3	-	2
Braille	1	-	-	-	1	1
Streamlined Mode	3	6	9	6	30	19
Text-to-Speech: Items	12	21	10	60	56	17
Text-to-Speech: Passages	1	-	-	-	-	-
Text-to-Speech: Passages & Items	711	-	-	-	-	-
Text-to-Speech: Stimuli	-	1	-	1	-	-
Text-to-Speech: Stimuli & Items	-	758	729	733	686	651
Non-Embedded Accommodations						
Alternate Response Options	-	1	4	1	3	1
Print-on-Demand: Stimuli and Items	5	10	3	11	11	7
Read-Aloud Stimuli	90	-	-	-	-	-
Scribe Items (Writing)	24	-	-	-	-	-
Speech-to-Text	58	90	91	113	93	106
Calculator	-	60	90	286	361	439
100s Number Table	-	195	142	108	42	13
Multiplication Table	-	276	385	487	486	432
Abacus	-	2	1	-	-	-

New Hampshire (adopted from NHSAS Technical Report, Vol 1, p. 9 - 10)

Accommodations	Grade					
	3	4	5	6	7	8
Embedded Accommodations						
Braille	-	-	-	-	-	1
Emboss: Stimuli & Items	-	-	-	-	-	1
Streamlined Mode	51	59	50	46	23	54
Text-to-Speech	1,654	1,831	1,557	1,558	1,268	1,191
Non-Embedded Accommodations						
100s Number Table	37	218	220	215	195	149
Abacus	9	2	3	4	2	12

Accommodations	Grade					
	3	4	5	6	7	8
ASL Human Signer	3	4	-	-	-	-
Print-on-Request	29	43	22	21	18	14
Read Aloud: Stimuli	376	490	379	265	186	155
Scribe	280	380	242	196	145	120
Speech-to-Text	119	160	152	95	95	68

West Virginia (adopted from WVGSA Technical Report, Vol 1, p. 10)

Accommodations	G3	G4	G5	G6	G7	G8
Embedded Accommodations						
Braille	3	3	2	2	2	2
Color Choices	1	3	2	7	0	1
Mouse Pointer	1	0	0	1	0	0
Streamlined Mode	35	30	18	24	22	24
Text-to-Speech: Instructions, Passages & Items	2,456	2,791	2,943	2,739	2,540	2,463
Non-Embedded Accommodations						
Print-on-Demand: Stimuli & Items	4	-	1	1	1	1

Wyoming (adopted from WY-TOPP Technical Report, Vol 1, p. 10)

Accommodations	Grade					
	3	4	5	6	7	8
Embedded Accommodations						
American Sign Language	1	-	-	1	1	-
Braille	-	-	-	1	-	1
Print-on-demand	36	22	20	17	38	23
Streamlined Mode	10	6	2	2	4	3
TTS: Items	685	831	757	662	618	579
TTS Tracking	459	513	466	367	411	381
Zoom	27	37	15	19	7	15
Non-Embedded Accommodations						
Calculator	-	-	-	6,808	7,359	7,238

Appendix B

DIF Analysis Results - ELA

Grade	Accommodation Type	Average number of student responses per item	Number of Items flagged						Percent of C level items (of total flagged items)
			DIF Category						
			A+	A-	B+	B-	C+	C-	
3	Text-to-Speech	365	37	56	0	7	0	0	0
	Non-Embedded Accommodations	65							
4	Text-to-Speech	409	23	47	1	3	0	0	0
	Non-Embedded Accommodations	81							
5	Text-to-Speech	501	23	51	0	4	0	0	0
	Non-Embedded Accommodations	84							
6	Text-to-Speech	330	22	59	0	0	0	0	0
	Non-Embedded Accommodations	57							
7	Text-to-Speech	425	21	49	0	2	0	1	1.37
	Non-Embedded Accommodations	58							
8	Text-to-Speech	342	23	34	0	2	0	0	0
	Non-Embedded Accommodations	53							

DIF Analysis Results - Mathematics

Grade	Accommodation type	Average number of student responses per item	Number of Items flagged						Percent of C level items (of total items flagged)
			DIF Category						
			A+	A-	B+	B-	C+	C-	
3	Text-to-Speech	368	17	61	0	10	0	5	5.37
	Calculator	NA							
	Non-Embedded Accommodations	45							
4	Text-to-Speech	374	21	71	0	7	0	7	6.6
	Calculator	4							
	Non-Embedded Accommodations	69							
5	Text-to-Speech	443	17	55	0	7	0	1	1.25
	Calculator	7							
	Non-Embedded Accommodations	79							
6	Text-to-Speech	488	12	26	0	0	0	0	0
	Calculator	637	17	5	0	1	0	0	0
	Non-Embedded Accommodations	88							
7	Text-to-Speech	414	13	45	0	0	0	0	0
	Calculator	29							
	Non-Embedded Accommodations	76							
8	Text-to-Speech	312	28	53	1	4	0	1	1.15
	Calculator	30							
	Non-Embedded Accommodations	55							

The Impact of the Pandemic on West Virginia Students

Introduction

Over the last year, K-12 teachers and students have experienced unprecedented educational challenges with ongoing school closures due to the pandemic. The resulting disruptions in student learning have led to significant concerns across states regarding student performance on the spring 2021 statewide educational assessments, with fears that scores will be compromised due to the lack of instructional coverage of content, differences in test modalities (remote versus onsite testing), and barriers to learning, such as a lack of internet access and varying levels of instructional and emotional support at home (Boyer, Dadey, & King, 2020). The CCSSO (2020a) has made several suggestions for ways in which the potential impact of the pandemic on learning can be investigated, such as the comparison of item statistics to previous years' test administrations and extensions of regularly performed analyses such as descriptive statistics and test reliability.

The WVGSA test window opened on March 29, 2021. Regardless of the circumstances, CAI psychometricians immediately begin to monitor item and overall assessment performance once the test becomes available to students. With the concerns over the changes in learning since the onset of the pandemic, in addition to their usual monitoring procedures, the CAI psychometric team began to incorporate additional analyses into their processes to assist WVDE in the identification of any changes in performance since the pre-pandemic assessment in spring 2019.

Changes in students' achievement levels and scale score, item statistics and test reliabilities were regularly monitored and reported as described in Sections 1 through 4. This early warning had the potential to provide WVDE with additional time in their efforts to form plans for remediation and support if warranted. Once the test window closed, additional analyses were carried out including residual and regression analyses that could provide further evidence for any changes. Prior to analyses, the samples were checked with respect (as available) to student disability, gender, ethnicity, socioeconomic status, and English learner representation to ensure they were consistent with the overall student population (Tables 1-5).

Specific research questions were as follows:

- With the changes in instruction introduced by the pandemic, did student academic performance change between 2019 and 2021 as evidenced by changes in overall scale scores, performance levels, correlational relationships between variables, and residuals?
- Have the test reliability coefficients and item parameters obtained during the current test administration changed since the spring 2019 administration?

Table 1: West Virginia Spring 2021 Math Student Gender Distributions

Grade	Gender	N_2019	Percent_2019	N_2021	Percentage_2021	Percent_Diff
3	Female	9058	49%	8020	49%	0%
3	Male	9502	51%	8328	51%	0%
4	Female	9473	49%	8073	49%	0%
4	Male	9732	51%	8553	51%	0%
5	Female	9482	48%	8221	49%	1%
5	Male	10075	52%	8596	51%	-1%
6	Female	9479	48%	8339	49%	1%
6	Male	10109	52%	8625	51%	-1%
7	Female	9129	48%	8366	48%	0%
7	Male	9720	52%	8983	52%	0%
8	Female	9088	48%	8333	48%	0%
8	Male	9968	52%	9035	52%	0%

Table 2: West Virginia Spring 2021 ELA Student Gender Distributions

Grade	Gender	N_2019	Percent_2019	N_2021	Percentage_2021	Percent_Diff
3	Female	9060	49%	8010	49%	0%
3	Male	9499	51%	8326	51%	0%
4	Female	9479	49%	8073	49%	0%
4	Male	9736	51%	8543	51%	0%
5	Female	9482	48%	8201	49%	1%
5	Male	10073	52%	8559	51%	-1%
6	Female	9486	48%	8300	49%	1%
6	Male	10116	52%	8569	51%	-1%
7	Female	9124	48%	8320	48%	0%
7	Male	9714	52%	8931	52%	0%
8	Female	9095	48%	8267	48%	0%
8	Male	9959	52%	8976	52%	0%

Table 3: West Virginia Spring 2021 Math Student Ethnicity Distributions

Grade	Ethnicity	N_2019	Percent_2019	N_2021	Percentage_2021	Percent_Diff
3	African American	714	4%	586	4%	0%
3	American Indian	8	0%	8	0%	0%
3	Asian	111	1%	85	1%	0%
3	Hispanic	377	2%	305	2%	0%
3	Multi-ethnic	751	4%	726	4%	0%
3	Pacific Islander	11	0%	7	0%	0%
3	White	16577	89%	14608	89%	0%
4	African American	783	4%	615	4%	0%
4	American Indian	22	0%	10	0%	0%
4	Asian	122	1%	107	1%	0%
4	Hispanic	377	2%	317	2%	0%
4	Multi-ethnic	769	4%	698	4%	0%
4	Pacific Islander	7	0%	5	0%	0%
4	White	17125	89%	14863	89%	0%
5	African American	734	4%	642	4%	0%
5	American Indian	19	0%	9	0%	0%
5	Asian	130	1%	95	1%	0%
5	Hispanic	400	2%	352	2%	0%
5	Multi-ethnic	754	4%	689	4%	0%
5	Pacific Islander	12	0%	8	0%	0%
5	White	17508	90%	15019	89%	-1%
6	African American	794	4%	676	4%	0%
6	American Indian	11	0%	25	0%	0%
6	Asian	137	1%	92	1%	0%
6	Hispanic	379	2%	323	2%	0%
6	Multi-ethnic	747	4%	628	4%	0%
6	Pacific Islander	9	0%	3	0%	0%
6	White	17509	89%	15216	90%	1%
7	African American	825	4%	638	4%	0%
7	American Indian	22	0%	19	0%	0%
7	Asian	120	1%	101	1%	0%
7	Hispanic	366	2%	352	2%	0%
7	Multi-ethnic	611	3%	629	4%	1%
7	Pacific Islander	12	0%	8	0%	0%
7	White	16893	90%	15602	90%	0%
8	African American	836	4%	645	4%	0%
8	American Indian	16	0%	10	0%	0%
8	Asian	137	1%	125	1%	0%
8	Hispanic	376	2%	319	2%	0%
8	Multi-ethnic	529	3%	614	4%	1%
8	Pacific Islander	8	0%	6	0%	0%
8	White	17154	90%	15649	90%	0%

Table 4: West Virginia Spring 2021 ELA Student Ethnicity Distributions

Grade	Ethnicity	N_2019	Percent_2019	N_2021	Percentage_2021	Percent_Diff
3	African American	714	4%	582	4%	0%
3	American Indian	8	0%	8	0%	0%
3	Asian	110	1%	86	1%	0%
3	Hispanic	380	2%	309	2%	0%
3	Multi-ethnic	750	4%	723	4%	0%
3	Pacific Islander	11	0%	7	0%	0%
3	White	16575	89%	14599	89%	0%
4	African American	783	4%	614	4%	0%
4	American Indian	22	0%	10	0%	0%
4	Asian	122	1%	107	1%	0%
4	Hispanic	381	2%	319	2%	0%
4	Multi-ethnic	770	4%	697	4%	0%
4	Pacific Islander	7	0%	6	0%	0%
4	White	17130	89%	14852	89%	0%
5	African American	732	4%	643	4%	0%
5	American Indian	19	0%	9	0%	0%
5	Asian	129	1%	95	1%	0%
5	Hispanic	411	2%	355	2%	0%
5	Multi-ethnic	753	4%	691	4%	0%
5	Pacific Islander	12	0%	8	0%	0%
5	White	17499	89%	14956	89%	0%
6	African American	790	4%	675	4%	0%
6	American Indian	11	0%	23	0%	0%
6	Asian	137	1%	92	1%	0%
6	Hispanic	389	2%	327	2%	0%
6	Multi-ethnic	745	4%	622	4%	0%
6	Pacific Islander	9	0%	3	0%	0%
6	White	17519	89%	15126	90%	1%
7	African American	825	4%	626	4%	0%
7	American Indian	22	0%	19	0%	0%
7	Asian	120	1%	101	1%	0%
7	Hispanic	371	2%	362	2%	0%
7	Multi-ethnic	610	3%	627	4%	1%
7	Pacific Islander	12	0%	7	0%	0%
7	White	16878	90%	15509	90%	0%
8	African American	835	4%	636	4%	0%
8	American Indian	16	0%	10	0%	0%
8	Asian	137	1%	124	1%	0%
8	Hispanic	385	2%	329	2%	0%
8	Multi-ethnic	529	3%	609	4%	1%
8	Pacific Islander	8	0%	6	0%	0%
8	White	17144	90%	15529	90%	0%

Table 5: West Virginia Spring 2021 Student LEP Distributions

Subject	Grade	EL	N_2019	Percent_2019	N_2021	Percent_2021	Percent_Diff
ELA	3	N	18402	99%	16233	99%	0%
ELA	3	Y	157	1%	119	1%	0%
ELA	4	N	19051	99%	16485	99%	0%
ELA	4	Y	164	1%	136	1%	0%
ELA	5	N	19407	99%	16674	99%	0%
ELA	5	Y	148	1%	94	1%	0%
ELA	6	N	19480	99%	16785	99%	0%
ELA	6	Y	122	1%	92	1%	0%
ELA	7	N	18731	99%	17159	99%	0%
ELA	7	Y	107	1%	108	1%	0%
ELA	8	N	18954	99%	17163	99%	0%
ELA	8	Y	100	1%	97	1%	0%
Math	3	N	18407	99%	16236	99%	0%
Math	3	Y	153	1%	115	1%	0%
Math	4	N	19046	99%	16506	99%	0%
Math	4	Y	159	1%	131	1%	0%
Math	5	N	19420	99%	16734	99%	0%
Math	5	Y	137	1%	90	1%	0%
Math	6	N	19475	99%	16890	99%	0%
Math	6	Y	113	1%	85	1%	0%
Math	7	N	18746	99%	17266	99%	0%
Math	7	Y	103	1%	93	1%	0%
Math	8	N	18965	100%	17292	99%	-1%
Math	8	Y	91	0%	88	1%	1%

1. Changes in Achievement Levels

A critical indicator of student academic performance on the state's content standards is the classification of student performance by achievement levels defined by cut scores established during a standard setting meeting conducted by CAI in 2018. Although the use of this data for accountability purposes has been suspended for 2021, it can serve as a valuable source of information for states to better understand the overall impact of the pandemic on student learning (CCSSO, 2020b). Tables 6 - 8 demonstrate the differences observed between spring 2021 and spring 2019 proficiency levels.

Table 6: West Virginia Math Student Achievement

Grade	AL	2018 Percentage	2019 Percentage	2021 Percentage	AL_Impact
3	1	23%	21%	32%	11%
3	2	29%	29%	29%	0%
3	3	27%	26%	23%	-3%
3	4	21%	25%	16%	-9%
4	1	22%	20%	31%	11%
4	2	33%	33%	35%	2%
4	3	22%	22%	18%	-4%
4	4	23%	25%	16%	-9%
5	1	29%	26%	38%	12%
5	2	31%	34%	33%	-1%
5	3	20%	20%	16%	-4%
5	4	20%	20%	13%	-7%
6	1	34%	34%	46%	12%
6	2	33%	32%	33%	1%
6	3	19%	19%	14%	-5%
6	4	14%	15%	7%	-8%
7	1	33%	34%	42%	8%
7	2	31%	30%	32%	2%
7	3	21%	19%	16%	-3%
7	4	14%	17%	10%	-7%
8	1	34%	33%	45%	12%
8	2	34%	31%	31%	0%
8	3	14%	15%	12%	-3%
8	4	17%	21%	12%	-9%

Table 7: West Virginia ELA Student Achievement Impact

Grade	AL	2018 Percentage	2019 Percentage	2021 Percentage	AL_Impact
3	1	21%	25%	35%	10%
3	2	32%	32%	31%	-1%
3	3	29%	26%	22%	-4%
3	4	18%	17%	12%	-5%
4	1	26%	24%	33%	9%
4	2	29%	28%	30%	2%
4	3	24%	25%	21%	-4%
4	4	20%	23%	16%	-7%
5	1	27%	26%	32%	6%
5	2	29%	27%	28%	1%
5	3	27%	26%	24%	-2%
5	4	17%	21%	16%	-5%
6	1	27%	25%	27%	2%
6	2	30%	31%	34%	3%
6	3	29%	31%	28%	-3%
6	4	14%	14%	11%	-3%
7	1	26%	27%	30%	3%
7	2	31%	31%	32%	1%
7	3	29%	29%	26%	-3%
7	4	15%	13%	12%	-1%
8	1	27%	25%	27%	2%
8	2	32%	32%	30%	-2%
8	3	28%	29%	28%	-1%
8	4	14%	15%	15%	0%

Table 8: West Virginia Student Proficiency Level Impact

Subject	Grade	2018_Proficient+	2019_Proficient+	2021_Proficient+	Impact_Proficiency
ELA	3	47%	43%	33%	-10%
ELA	4	44%	48%	37%	-11%
ELA	5	44%	47%	40%	-7%
ELA	6	43%	45%	39%	-6%
ELA	7	44%	42%	38%	-4%
ELA	8	42%	44%	43%	-1%
Math	3	48%	51%	38%	-13%
Math	4	45%	47%	34%	-13%
Math	5	40%	40%	28%	-12%
Math	6	33%	34%	21%	-13%
Math	7	35%	36%	26%	-10%
Math	8	31%	36%	24%	-12%

2. Changes in Scale Scores

The mean scale scores provide a point estimate that can be compared to previous years' mean scores, while the standard deviation provides us with a measure of the overall spread of scores from the mean. For example, if scores have potentially been impacted by the pandemic, we might expect to see some significant downward shifts in the mean scores and/or larger standard deviations when compared to the 2019 scores (although additional analyses would be required to determine if such differences were statistically significant). Psychometricians began monitoring these values on a weekly basis (Table 9). The WVDE was able to obtain and monitor these statistics upon demand with the use of the CRS system.

Variability between different test administrations is always expected regardless of the circumstances. Therefore, in an effort to reduce some of this variability, it was also useful to compare spring 2021 scores to the overall mean scale scores of both spring 2018 and spring 2019 combined (Table 10).

Table 9: West Virginia Spring 2021 Scale Score Impact

Subject	Grade	2018_N	2018_Mean_SS	2018_SD	2019_N	2019_Mean_SS	2019_SD	2021_N	2021_Mean_SS	2021_SD	Impact_SS
ELA	3	19343	580	39.10	18559	577	40.03	16352	566	41	-11
ELA	4	19751	591	43.84	19215	594	46.02	16621	583	46	-11
ELA	5	19874	612	44.30	19555	617	46.03	16768	608	45	-9
ELA	6	19085	627	48.21	19602	629	47.89	16877	624	46	-5
ELA	7	19280	633	49.60	18838	631	48.50	17267	627	49	-4
ELA	8	19374	644	48.84	19054	646	49.49	17260	644	51	-2
Math	3	19409	422	34.18	18560	425	33.74	16351	414	36	-11
Math	4	19790	449	41.72	19205	450	41.27	16637	438	42	-12
Math	5	19934	472	52.07	19557	474	48.24	16824	459	49	-15
Math	6	19125	493	55.01	19588	492	57.06	16975	477	53	-15
Math	7	19324	523	59.95	18849	523	64.42	17359	510	60	-13
Math	8	19444	552	72.24	19056	555	76.90	17380	534	73	-21

Table 10: West Virginia Spring 2021 Scale Score Impact

Subject	Grade	2018_Mean_SS	2019_Mean_SS	Mean_18_19_SS	2021_Mean_SS	Mean_Diff
ELA	3	580	577	578	566	-12
ELA	4	591	594	592	583	-10
ELA	5	612	617	614	608	-6
ELA	6	627	629	628	624	-4
ELA	7	633	631	632	627	-5
ELA	8	644	646	645	644	-1
Math	3	422	425	424	414	-10
Math	4	449	450	450	438	-12
Math	5	472	474	473	459	-14
Math	6	493	492	492	477	-16
Math	7	523	523	523	510	-13
Math	8	552	555	554	534	-20

3. Changes in Test Reliability

Test reliability provides information regarding how consistently a test is measuring the construct of interest. Theoretically, if a student was tested twice using an assessment with a high reliability coefficient, we would expect that the student would receive a similar score reflective of their true ability level with both test administrations. However, if the same student tested the second time while feeling extremely ill, it is unlikely that a similar score would be produced due to the variation in measurement introduced by the illness (Haertel, 2006). Such variations are referred to as measurement error, and when a high degree of error is introduced from external influences, test reliability coefficients can be negatively impacted.

With this, test reliability for the math and ELA tests began to be computed once 30% of the student population had been tested and once weekly thereafter to monitor any changes in the reliability coefficients. If a significant decrease in reliability had been observed, this would have suggested that additional measurement error may have been introduced due to the changes in either student learning or the testing conditions (Table 11).

Table 11: West Virginia Spring 2021 Test Reliability Impact

Subject	Grade	2018_N	2018_Reliability	2019_N	2019_Reliability	2021_N	2021_Reliability	Impact_Rel
ELA	3	19343	0.89	18559	0.89	16352	0.89	0.00
ELA	4	19751	0.88	19215	0.88	16621	0.87	-0.01
ELA	5	19874	0.89	19555	0.90	16768	0.89	-0.01
ELA	6	19085	0.90	19602	0.89	16877	0.89	0.00
ELA	7	19280	0.88	18838	0.90	17267	0.90	0.00
ELA	8	19374	0.90	19054	0.91	17260	0.91	0.00
Math	3	19409	0.92	18560	0.92	16351	0.91	-0.01
Math	4	19790	0.92	19205	0.92	16637	0.91	-0.01
Math	5	19934	0.93	19557	0.92	16824	0.89	-0.03
Math	6	19125	0.92	19588	0.89	16975	0.88	-0.01
Math	7	19324	0.93	18849	0.87	17359	0.87	0.00
Math	8	19444	0.91	19056	0.89	17380	0.88	-0.01

4. Changes in Item Statistics

Item statistics are regularly monitored by CAI psychometricians throughout the testing window. Item difficulty (P-value) is the average item scores across all students with lower numbers indicating higher difficulty levels. Significant increases in item difficulty would have suggested that additional investigations may be warranted (Tables 12-13). Items that presented a decrease in p-value greater than 0.1 were flagged, and the percentage of the flagged items within each content standard was computed. The items that were administered both in spring 2019 and 2021 (common items in the pool) to more than 100 students (items compared) were included in this p-value evaluation.

Table 12: West Virginia Spring 2019 - 2021 Item PValues Math

Grade	ContentLevelID	N_Common_Items	N_compared	N_Flagged	Prop_Flagged
3	MDG 3.G.c1 M.3.25	12	8	4	50%
3	MDG 3.MD.c1 M.3.16	8	8	2	25%
3	MDG 3.MD.c2 M.3.18	14	14	4	29%
3	MDG 3.MD.c3 M.3.22 M.3.22b	5	5	2	40%
3	MDG 3.MD.c3 M.3.22 M.3.22d	4	2	1	50%
3	MDG 3.MD.c4 M.3.23	8	7	4	57%
3	NBTF 3.NBT.c1 M.3.10	28	28	8	29%
3	NBTF 3.NBT.c1 M.3.11	31	31	5	16%
3	NBTF 3.NBT.c1 M.3.12	35	35	7	20%
3	NBTF 3.NF.c1 M.3.13	48	42	8	19%
3	NBTF 3.NF.c1 M.3.14 M.3.14a	15	12	3	25%
3	NBTF 3.NF.c1 M.3.14 M.3.14b	44	32	5	16%
3	NBTF 3.NF.c1 M.3.15 M.3.15b	18	16	3	19%
3	NBTF 3.NF.c1 M.3.15 M.3.15d	9	9	1	11%
3	OAT 3.OAT.c1 M.3.1	9	9	2	22%
3	OAT 3.OAT.c1 M.3.2	6	6	1	17%
3	OAT 3.OAT.c1 M.3.3	14	13	3	23%
3	OAT 3.OAT.c1 M.3.4	9	8	3	38%
3	OAT 3.OAT.c2 M.3.5	22	14	1	7%
3	OAT 3.OAT.c2 M.3.6	20	18	2	11%
3	OAT 3.OAT.c3 M.3.7	15	6	5	83%
3	OAT 3.OAT.c4 M.3.8	25	19	4	21%
3	OAT 3.OAT.c4 M.3.9	18	12	4	33%
4	MDG 4.G.c1 M.4.27	7	2	1	50%
4	MDG 4.G.c1 M.4.28	9	9	2	22%
4	MDG 4.MD.c1 M.4.20	12	12	1	8%
4	MDG 4.MD.c1 M.4.21	9	8	1	12%
4	MDG 4.MD.c2 M.4.22	8	7	2	29%
4	MDG 4.MD.c3 M.4.23 M.4.23a	3	2	2	100%
4	MDG 4.MD.c3 M.4.23 M.4.23b	4	4	1	25%
4	MDG 4.MD.c3 M.4.24	10	10	1	10%
4	MDG 4.MD.c3 M.4.25	6	6	2	33%
4	NBTF 4.NBT.c2 M.4.10	21	21	8	38%
4	NBTF 4.NBT.c2 M.4.11	19	19	9	47%
4	NBTF 4.NF.c1 M.4.12	23	21	5	24%
4	NBTF 4.NF.c1 M.4.13	12	11	2	18%
4	NBTF 4.NF.c2 M.4.14 M.4.14a	12	12	4	33%
4	NBTF 4.NF.c2 M.4.14 M.4.14c	6	6	2	33%
4	NBTF 4.NF.c2 M.4.15 M.4.15a	4	4	1	25%
4	NBTF 4.NF.c3 M.4.16	21	20	1	5%

4	NBTF 4.NF.c3 M.4.17	18	12	1	8%
4	NBTF 4.NF.c3 M.4.18	15	15	4	27%
4	OAT 4.OAT.c1 M.4.1	23	21	3	14%
4	OAT 4.OAT.c1 M.4.2	14	14	3	21%
4	OAT 4.OAT.c2 M.4.4	26	23	4	17%
4	OAT 4.OAT.c3 M.4.5	21	16	2	12%
5	MDG 5.G.c1 M.5.24	8	8	2	25%
5	MDG 5.G.c2 M.5.25	6	6	1	17%
5	MDG 5.G.c2 M.5.26	5	4	2	50%
5	MDG 5.MD.c1 M.5.18	16	11	2	18%
5	MDG 5.MD.c2 M.5.19	14	14	4	29%
5	MDG 5.MD.c3 M.5.20 M.5.20a	3	2	2	100%
5	MDG 5.MD.c3 M.5.21	6	4	1	25%
5	MDG 5.MD.c3 M.5.22 M.5.22b	10	10	1	10%
5	NBTF 5.NBT.c1 M.5.4-5 M.5.5	18	7	3	43%
5	NBTF 5.NBT.c1 M.5.6 M.5.6a	10	2	1	50%
5	NBTF 5.NBT.c1 M.5.7	18	18	4	22%
5	NBTF 5.NBT.c2 M.5.10	20	20	4	20%
5	NBTF 5.NBT.c2 M.5.8	23	23	9	39%
5	NBTF 5.NBT.c2 M.5.9	24	24	5	21%
5	NBTF 5.NF.c1 M.5.11	25	25	4	16%
5	NBTF 5.NF.c1 M.5.12	14	14	3	21%
5	NBTF 5.NF.c2 M.5.13	14	14	2	14%
5	NBTF 5.NF.c2 M.5.14 M.5.14a	8	6	2	33%
5	NBTF 5.NF.c2 M.5.15 M.5.15a	8	3	3	100%
5	NBTF 5.NF.c2 M.5.16	24	19	2	11%
5	NBTF 5.NF.c2 M.5.17 M.5.17a	2	1	1	100%
5	NBTF 5.NF.c2 M.5.17 M.5.17c	5	2	2	100%
5	OAT 5.OAT.c1 M.5.1	32	32	4	12%
5	OAT 5.OAT.c1 M.5.2	20	18	2	11%
5	OAT 5.OAT.c2 M.5.3	16	16	3	19%
6	EE 6.EE.c1 M.6.12	14	13	2	15%
6	EE 6.EE.c1 M.6.13 M.6.13b	4	4	1	25%
6	EE 6.EE.c2 M.6.16	9	9	2	22%
6	EE 6.EE.c2 M.6.18	23	21	2	10%
6	EE 6.EE.c2 M.6.19	16	14	3	21%
6	EE 6.EE.c3 M.6.20	28	15	2	13%
6	GSP 6.G.c1 M.6.22	6	6	1	17%
6	GSP 6.G.c1 M.6.23	8	8	3	38%
6	GSP 6.SP.c1 M.6.26	8	7	1	14%
6	GSP 6.SP.c1 M.6.27	2	2	1	50%
6	GSP 6.SP.c2 M.6.28	5	5	1	20%
6	RPNS 6.NS.c1 M.6.4	24	18	4	22%
6	RPNS 6.NS.c2 M.6.5	10	10	3	30%
6	RPNS 6.NS.c2 M.6.7	11	8	1	12%
6	RPNS 6.NS.c3 M.6.9 M.6.9a	3	3	1	33%
6	RPNS 6.RP.c1 M.6.1	50	38	10	26%
6	RPNS 6.RP.c1 M.6.2	27	21	1	5%
6	RPNS 6.RP.c1 M.6.3 M.6.3a	11	4	2	50%
6	RPNS 6.RP.c1 M.6.3 M.6.3c	6	1	1	100%
7	EE 7.EE.c1 M.7.7	19	13	1	8%
7	EE 7.EE.c2 M.7.10 M.7.10a	10	5	1	20%
7	EE 7.EE.c2 M.7.9	15	15	1	7%
7	G 7.G.c1 M.7.11	9	9	3	33%

7	G 7.G.c1 M.7.12	10	10	2	20%
7	G 7.G.c1 M.7.13	12	11	1	9%
7	G 7.G.c2 M.7.14	13	11	1	9%
7	G 7.G.c2 M.7.15	14	14	2	14%
7	RPNS 7.NS.c1 M.7.4 M.7.4b	13	10	2	20%
7	RPNS 7.NS.c1 M.7.4 M.7.4d	6	5	1	20%
7	RPNS 7.NS.c1 M.7.5 M.7.5c	14	6	2	33%
7	RPNS 7.NS.c1 M.7.6	20	20	1	5%
7	RPNS 7.RP.c1 M.7.2 M.7.2c	5	3	1	33%
7	RPNS 7.RP.c1 M.7.3	31	24	4	17%
7	SP 7.SP.c1 M.7.17	1	1	1	100%
7	SP 7.SP.c1 M.7.18	7	4	1	25%
7	SP 7.SP.c3 M.7.23	8	8	2	25%
8	EENS 8.EE.c1 M.8.4	9	8	2	25%
8	EENS 8.EE.c1 M.8.5	10	8	1	12%
8	EENS 8.EE.c1 M.8.6	4	4	1	25%
8	EENS 8.EE.c2 M.8.8	11	7	1	14%
8	EENS 8.EE.c3 M.8.9 M.8.9a	11	8	1	12%
8	EENS 8.NS.c1 M.8.2	26	24	5	21%
8	F 8.F.c1 M.8.11	16	13	1	8%
8	F 8.F.c2 M.8.14	15	15	1	7%
8	F 8.F.c2 M.8.15	18	14	1	7%
8	GSP 8.G.c1 M.8.16 M.8.16a	5	5	2	40%
8	GSP 8.G.c1 M.8.17	9	9	3	33%
8	GSP 8.G.c1 M.8.19	5	5	2	40%
8	GSP 8.G.c2 M.8.21	11	7	2	29%
8	GSP 8.G.c2 M.8.22	14	12	1	8%
8	GSP 8.G.c3 M.8.24	24	11	3	27%
8	GSP 8.SP.c1 M.8.25	11	10	3	30%
8	GSP 8.SP.c1 M.8.26	16	16	3	19%

Table 13: West Virginia Spring 2019 - 2021 Item PValues ELA

Grade	ContentLevelID	N_Common_Items	N_compared	N_Flagged	Prop_Flagged
3	IT 3.CS ELA.3.10	14	14	1	7%
3	IT 3.IKI ELA.3.15	10	8	1	12%
3	L 3.CSE ELA.3.36d	3	3	1	33%
3	L 3.CSE ELA.3.36e	6	6	1	17%
3	L 3.CSE ELA.3.36g	6	6	2	33%
3	L 3.CSE ELA.3.36h	6	6	2	33%
3	L 3.CSE ELA.3.36i	5	5	1	20%
3	L 3.CSE ELA.3.37c	2	2	1	50%
3	L 3.CSE ELA.3.37d	4	4	1	25%
3	L 3.CSE ELA.3.37e	6	6	3	50%
3	L 3.CSE ELA.3.37f	3	3	1	33%
3	LT 3.KID ELA.3.3	18	18	1	6%
4	IT 4.CS ELA.4.10	22	21	2	10%
4	IT 4.IKI ELA.4.16	12	8	1	12%
4	IT 4.IKI ELA.4.17	7	7	1	14%
4	IT 4.KID ELA.4.4	21	19	2	11%
4	L 4.CSE ELA.4.36b	1	1	1	100%
4	L 4.CSE ELA.4.36c	7	7	1	14%
4	L 4.CSE ELA.4.36e	10	10	1	10%
4	L 4.CSE ELA.4.36g	4	4	2	50%
4	L 4.CSE ELA.4.37a	9	9	2	22%
4	L 4.CSE ELA.4.37b	1	1	1	100%
4	L 4.CSE ELA.4.37c	7	7	1	14%
4	L 4.CSE ELA.4.37d	6	6	2	33%
4	LT 4.CS ELA.4.7	14	14	1	7%
4	LT 4.KID ELA.4.3	17	17	1	6%
5	IT 5.CS ELA.5.12	10	10	1	10%
5	IT 5.IKI ELA.5.16	18	15	2	13%
5	IT 5.IKI ELA.5.17	8	7	1	14%
5	IT 5.KID ELA.5.4	14	14	2	14%
5	L 5.CSE ELA.5.36b	5	5	1	20%
5	L 5.CSE ELA.5.36d	10	10	3	30%
5	L 5.CSE ELA.5.36e	6	6	1	17%
5	L 5.CSE ELA.5.37c	3	3	1	33%
5	L 5.CSE ELA.5.37e	6	6	1	17%
5	SL 5.CaC ELA.5.32	3	3	2	67%
5	W 5.TTP ELA.5.21a	1	4	3	75%
6	IT 6.IKI ELA.6.15	7	5	1	20%
6	IT 6.IKI ELA.6.16	17	12	1	8%
6	IT 6.KID ELA.6.4	28	26	2	8%
6	L 6.CSE ELA.6.37a	7	7	1	14%
6	L 6.CSE ELA.6.37b	6	6	1	17%
6	LT 6.CS ELA.6.7	18	17	2	12%
6	LT 6.KID ELA.6.1	18	17	1	6%
6	LT 6.KID ELA.6.2	14	14	2	14%
7	IT 7.CS ELA.7.10	18	16	1	6%
7	IT 7.CS ELA.7.11	18	17	1	6%
7	L 7.CSE ELA.7.36b	8	8	1	12%
7	L 7.CSE ELA.7.36c	14	14	2	14%
7	L 7.CSE ELA.7.37a	8	8	1	12%

7	L 7.CSE ELA.7.37b	11	11	4	36%
7	LT 7.CS ELA.7.7	15	15	1	7%
7	LT 7.KID ELA.7.3	21	20	1	5%
8	IT 8.CS ELA.8.11	13	12	1	8%
8	L 8.CSE ELA.8.37a	12	12	2	17%
8	L 8.CSE ELA.8.37c	7	7	2	29%
8	L 8.KL ELA.8.38b	14	14	1	7%
8	L 8.KL ELA.8.38c	6	6	1	17%
8	L 8.VAU ELA.8.39a	10	10	1	10%
8	LT 8.KID ELA.8.2	5	5	1	20%

5. Residual Analyses

In statistical models, a residual is the difference between the observed value and the expected value that the model predicts for that observation. If assessments fail to maintain the same level of score validity/stability between testing occasions, this may be manifested in the observed changes of the magnitudes and/or patterns of the residuals. For each item, a model-predicted (or expected) score was computed given the student's ability estimate and the item parameters known in the item bank using the IRT models adopted for the WVGS. The item-level residuals were summed for a student (across all items) and for an aggregated unit (across all items and students in a grade). The aggregated residuals for pre- and post-pandemic administrations were compared at the test-level.

For dichotomous items, the 3PL IRT model was used to compute the expected score (1), and for polytomous items, the generalized partial credit model was used (2).

$$E(z_{ij}) = c_i + (1 - c_i) \frac{\exp(D a_i (\theta_j - b_i))}{1 + \exp(D a_i (\theta_j - b_i))}, \quad (1)$$

where, $E(z_{ij})$ is the expected score for item i for student j , c_i is the pseudo-guessing parameter, a_i is the item discrimination parameter, b_i is the item location (difficulty) parameter, and D is a constant fixed at 1.7, bringing the logistic into coincidence with the probit model. Student estimated ability is represented by θ_j .

$$E(z_{ij}) = \sum_{l=1}^{m_i} \frac{l \exp(\sum_{k=1}^l D a_i (\theta_j - b_{i,k}))}{1 + \sum_{l=1}^{m_i} \exp(\sum_{k=1}^l D a_i (\theta_j - b_{i,k}))}, \quad (2)$$

where, $b_{i,k}$ are the step parameter values with the maximum possible score m_i . For each item i the residual between observed and expected score for each student is defined as follows:

$$\delta_{ij} = z_{ij} - E(z_{ij}), \quad (3)$$

A positive δ_{ij} indicates that student j obtains a higher score on item i than the expected score given his or her ability and item parameters while a negative δ_{ij} indicates that student j performs lower than expected on item i .

The residuals (δ_{ij}) are summed across all items that student j was given and the sum of the residuals is divided by the total points possible for the test form student j took. Residual at test level for each individual is defined as (4):

$$\delta_j = \frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n (\delta_{ij}), \quad (4)$$

where m_i is the points possible for item i and n is the total number of items in the test form.

The residual analyses were conducted for spring 2019 WVGSA data to establish the reference values against which spring 2021 results were compared.

To investigate the changes in residuals at different ability levels, students were grouped into ten groups based on their estimated θ value. Within the lowest and highest theta boundaries for each grade, theta scales were divided into 10 bins. The bins with a lower number indicate a lower ability group and the groups with a higher number indicate a higher ability group.

The test-level residuals (4) were averaged across all students within a bin and the standard error (se) was computed.

$$\delta_g = \frac{1}{n_g} \sum_{j=1}^{n_g} (\delta_j), \quad (5)$$

$$se(\delta_g) = sd(\delta_g) / \sqrt{n_g}, \quad (6)$$

where g is the ability group, n_g is the number of students in group g , and the $sd(\delta_g)$ is the standard deviation of the test-level residual within g .

Note that there were no or few students in the higher bins for some tests. In the results section, the statistics calculated for those small groups were retained in the tables and plots to show the student's ability level distribution. However, the large standard error (6) associated with those small groups should be considered and caution is needed when interpreting the meaning of those values.

Figures 1 and 2 show the mean residual plot for ELA and mathematics, respectively. The means of residuals in each ability group (5) are marked by grade. As can be seen in the residual plots, there are significant outliers in some of the higher ability groups. However, Table 14 and 15 show that all of the outlier values are based on a very small number of students. The standard deviation (SD) and the number of students (N) are included in Table 15 and Table 16 and the lower and upper limit of the 95% Confidence Interval (CI) for the spring 21 mean residuals are presented in the tables to assist with the interpretation of those values.

Table 14: Math Residuals

Grade	Group	2019_N	2019_Mean	2019_SD	2021_N	2021_Mean	2021_SD	CI_LL_Sp21	CI_UL_Sp21
3	1	263	-0.0037	0.0223	427	-0.0033	0.0214	-0.0053	-0.0013
3	2	352	0.0008	0.0202	588	0.0006	0.0197	-0.0010	0.0022
3	3	955	-0.0028	0.0218	1405	-0.0001	0.0217	-0.0013	0.0011
3	4	2801	-0.0030	0.0218	3330	-0.0047	0.0240	-0.0055	-0.0039
3	5	5250	-0.0021	0.0212	4716	-0.0018	0.0223	-0.0024	-0.0012
3	6	5078	0.0000	0.0200	3764	-0.0013	0.0220	-0.0021	-0.0005
3	7	2931	0.0023	0.0195	1651	0.0017	0.0210	0.0007	0.0027
3	8	777	0.0041	0.0183	398	0.0035	0.0196	0.0015	0.0055
3	9	123	0.0074	0.0158	50	0.0075	0.0152	0.0032	0.0118
3	10	28	0.0082	0.0147	15	0.0093	0.0180	0.0003	0.0183
4	1	359	-0.0045	0.0207	457	-0.0061	0.0158	-0.0075	-0.0047
4	2	425	0.0002	0.0188	608	-0.0013	0.0164	-0.0027	0.0001
4	3	1141	-0.0021	0.0204	1729	-0.0060	0.0218	-0.0070	-0.0050
4	4	3250	-0.0057	0.0210	3680	-0.0073	0.0230	-0.0081	-0.0065
4	5	5886	-0.0046	0.0223	5209	-0.0050	0.0243	-0.0056	-0.0044
4	6	5074	-0.0013	0.0227	3305	-0.0007	0.0243	-0.0015	0.0001
4	7	2404	0.0000	0.0217	1299	0.0000	0.0226	-0.0012	0.0012
4	8	565	0.0037	0.0203	284	0.0026	0.0214	0.0001	0.0051
4	9	84	0.0030	0.0194	45	0.0052	0.0192	-0.0005	0.0109
4	10	17	0.0072	0.0183	5	0.0039	0.0220	-0.0155	0.0233
5	1	427	-0.0077	0.0248	625	-0.0049	0.0257	-0.0069	-0.0029
5	2	414	-0.0028	0.0217	572	-0.0032	0.0257	-0.0054	-0.0010
5	3	1317	-0.0092	0.0228	1680	-0.0062	0.0255	-0.0074	-0.0050
5	4	3985	-0.0051	0.0229	4533	-0.0036	0.0250	-0.0044	-0.0028
5	5	5957	-0.0010	0.0243	4862	0.0020	0.0253	0.0012	0.0028
5	6	4614	0.0023	0.0246	3087	0.0015	0.0257	0.0005	0.0025
5	7	2155	0.0038	0.0215	1167	0.0024	0.0234	0.0010	0.0038
5	8	588	0.0031	0.0197	256	0.0013	0.0194	-0.0011	0.0037
5	9	78	0.0022	0.0166	30	0.0021	0.0131	-0.0026	0.0068
5	10	19	0.0054	0.0184	5	0.0120	0.0216	-0.0068	0.0308
6	1	553	-0.0007	0.0237	564	0.0011	0.0255	-0.0011	0.0033
6	2	790	0.0002	0.0219	813	0.0003	0.0234	-0.0013	0.0019
6	3	2029	-0.0022	0.0218	2525	-0.0038	0.0231	-0.0048	-0.0028
6	4	4410	-0.0006	0.0212	5126	-0.0023	0.0223	-0.0029	-0.0017
6	5	5763	0.0004	0.0193	4797	-0.0009	0.0194	-0.0015	-0.0003
6	6	4138	-0.0002	0.0203	2335	-0.0009	0.0207	-0.0017	-0.0001
6	7	1569	0.0010	0.0233	669	-0.0003	0.0234	-0.0021	0.0015
6	8	291	0.0021	0.0232	118	0.0082	0.0252	0.0037	0.0127
6	9	35	0.0073	0.0148	17	0.0000	0.0204	-0.0098	0.0098
6	10	6	-0.0086	0.0095	0	NA	NA	NA	NA
7	1	757	0.0013	0.0330	597	-0.0067	0.0281	-0.0091	-0.0043
7	2	653	0.0055	0.0298	859	-0.0022	0.0279	-0.0042	-0.0002
7	3	1828	-0.0009	0.0309	2342	-0.0023	0.0297	-0.0035	-0.0011
7	4	3816	-0.0029	0.0294	4364	-0.0047	0.0293	-0.0055	-0.0039
7	5	5196	-0.0034	0.0283	4893	-0.0064	0.0278	-0.0072	-0.0056
7	6	4033	-0.0015	0.0271	2896	-0.0035	0.0274	-0.0045	-0.0025
7	7	1946	0.0037	0.0248	1108	0.0000	0.0263	-0.0016	0.0016
7	8	517	0.0090	0.0231	238	0.0053	0.0279	0.0018	0.0088
7	9	91	0.0107	0.0265	45	0.0125	0.0338	0.0027	0.0223
7	10	11	0.0297	0.0206	7	0.0442	0.0381	0.0160	0.0724
8	1	640	0.0040	0.0305	590	-0.0014	0.0268	-0.0036	0.0008
8	2	716	0.0023	0.0267	955	0.0008	0.0245	-0.0008	0.0024
8	3	1740	-0.0018	0.0238	2565	-0.0026	0.0249	-0.0036	-0.0016
8	4	3652	-0.0043	0.0236	4227	-0.0013	0.0246	-0.0021	-0.0005
8	5	5147	-0.0057	0.0239	4718	-0.0045	0.0265	-0.0053	-0.0037
8	6	4511	-0.0054	0.0244	2994	-0.0028	0.0270	-0.0038	-0.0018
8	7	2016	-0.0009	0.0279	1030	-0.0004	0.0272	-0.0020	0.0012
8	8	517	0.0043	0.0262	240	0.0051	0.0224	0.0024	0.0078
8	9	91	0.0050	0.0256	42	0.0095	0.0200	0.0034	0.0156
8	10	22	0.0089	0.0161	7	-0.0064	0.0102	-0.0140	0.0012

Table 15: ELA Residuals

Grade	Group	2019_N	2019_Mean	2019_SD	2021_N	2021_Mean	2021_SD	CI_LL_Sp21	CI_UL_Sp21
3	1	60	0.0076	0.0412	59	0.0104	0.0405	0.0000	0.0208
3	2	183	0.0069	0.0346	272	0.0091	0.0338	0.0052	0.0130
3	3	1245	0.0078	0.0347	1815	0.0112	0.0352	0.0096	0.0128
3	4	3464	0.0063	0.0356	4020	0.0126	0.0362	0.0114	0.0138
3	5	5490	-0.0037	0.0322	4702	0.0037	0.0341	0.0027	0.0047
3	6	5333	-0.0019	0.0262	3806	0.0007	0.0298	-0.0003	0.0017
3	7	2444	0.0031	0.0249	1425	0.0032	0.0263	0.0018	0.0046
3	8	311	0.0064	0.0232	228	0.0123	0.0238	0.0092	0.0154
3	9	21	0.0032	0.0281	9	-0.0015	0.0293	-0.0207	0.0177
3	10	2	0.0046	0.0032	0	NA	NA	NA	NA
4	1	150	0.0139	0.0355	139	0.0143	0.0357	0.0084	0.0202
4	2	408	0.0074	0.0303	484	0.0104	0.0321	0.0075	0.0133
4	3	1627	0.0035	0.0297	2025	0.0090	0.0312	0.0076	0.0104
4	4	3766	-0.0012	0.0283	4241	0.0042	0.0309	0.0032	0.0052
4	5	5623	-0.0045	0.0274	4809	0.0027	0.0307	0.0019	0.0035
4	6	5062	-0.0022	0.0256	3396	-0.0007	0.0307	-0.0017	0.0003
4	7	2204	0.0029	0.0244	1332	0.0025	0.0285	0.0009	0.0041
4	8	349	0.0090	0.0243	177	0.0045	0.0274	0.0004	0.0086
4	9	23	0.0109	0.0258	13	0.0137	0.0249	0.0002	0.0272
4	10	2	0.0250	0.0415	0	NA	NA	NA	NA
5	1	91	0.0064	0.0347	73	0.0087	0.0349	0.0007	0.0167
5	2	345	0.0080	0.0297	392	0.0117	0.0309	0.0086	0.0148
5	3	1547	0.0026	0.0292	1790	0.0097	0.0308	0.0083	0.0111
5	4	3631	-0.0032	0.0305	3773	0.0017	0.0314	0.0007	0.0027
5	5	5743	-0.0040	0.0283	4925	-0.0052	0.0293	-0.0060	-0.0044
5	6	4993	-0.0012	0.0243	3993	-0.0034	0.0274	-0.0042	-0.0026
5	7	2659	0.0037	0.0221	1590	-0.0009	0.0243	-0.0021	0.0003
5	8	509	0.0058	0.0221	216	0.0067	0.0215	0.0038	0.0096
5	9	31	0.0031	0.0213	8	-0.0022	0.0290	-0.0222	0.0178
5	10	4	0.0098	0.0179	0	NA	NA	NA	NA
6	1	166	0.0005	0.0344	115	0.0011	0.0319	-0.0048	0.0070
6	2	452	0.0024	0.0311	427	0.0042	0.0308	0.0013	0.0071
6	3	1837	-0.0002	0.0330	1792	0.0061	0.0313	0.0047	0.0075
6	4	3723	-0.0006	0.0332	3619	0.0021	0.0345	0.0009	0.0033
6	5	5532	-0.0080	0.0350	5075	-0.0058	0.0336	-0.0068	-0.0048
6	6	5205	-0.0065	0.0283	4046	-0.0047	0.0278	-0.0055	-0.0039
6	7	2239	0.0017	0.0277	1563	0.0004	0.0283	-0.0010	0.0018
6	8	416	0.0008	0.0259	221	0.0031	0.0267	-0.0004	0.0066
6	9	25	0.0058	0.0193	9	-0.0104	0.0227	-0.0253	0.0045
6	10	3	-0.0120	0.0036	2	-0.0062	0.0221	-0.0370	0.0246
7	1	121	0.0127	0.0335	126	0.0092	0.0331	0.0033	0.0151
7	2	472	0.0101	0.0341	576	0.0091	0.0327	0.0064	0.0118
7	3	1941	0.0091	0.0351	2001	0.0101	0.0338	0.0085	0.0117
7	4	3902	0.0062	0.0424	3796	0.0044	0.0397	0.0032	0.0056
7	5	5545	0.0011	0.0365	5060	0.0002	0.0325	-0.0008	0.0012
7	6	4729	0.0111	0.0303	3813	0.0015	0.0266	0.0007	0.0023
7	7	1824	0.0127	0.0303	1663	0.0030	0.0257	0.0018	0.0042
7	8	272	0.0097	0.0276	210	0.0094	0.0293	0.0055	0.0133
7	9	27	0.0025	0.0226	6	0.0131	0.0196	-0.0026	0.0288
7	10	0	NA	NA	0	NA	NA	NA	NA
8	1	46	0.0081	0.0275	64	0.0150	0.0307	0.0076	0.0224
8	2	246	0.0104	0.0303	272	0.0062	0.0326	0.0023	0.0101
8	3	1630	0.0082	0.0338	1527	0.0066	0.0333	0.0048	0.0084
8	4	3667	0.0010	0.0375	3493	-0.0034	0.0346	-0.0046	-0.0022
8	5	5844	-0.0051	0.0325	5104	-0.0108	0.0307	-0.0116	-0.0100
8	6	5067	0.0013	0.0284	4496	-0.0055	0.0275	-0.0063	-0.0047
8	7	2145	0.0055	0.0263	1953	-0.0005	0.0255	-0.0017	0.0007
8	8	371	0.0023	0.0258	315	0.0003	0.0228	-0.0022	0.0028
8	9	32	-0.0071	0.0121	18	-0.0118	0.0256	-0.0236	0.0000
8	10	4	-0.0042	0.0085	1	0.0091	NA	NA	NA

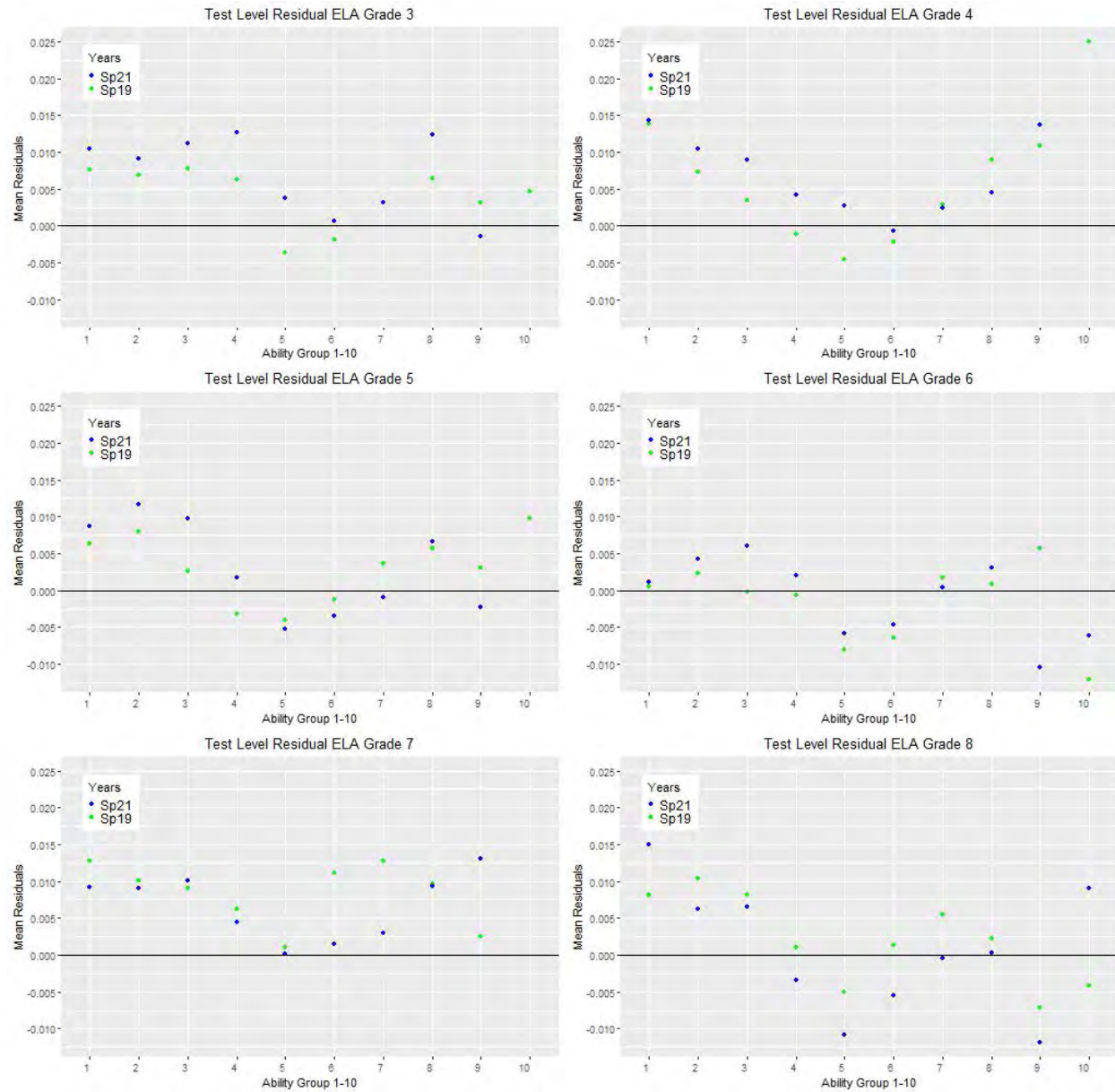


Figure 1: ELA Test Level Residual Comparison

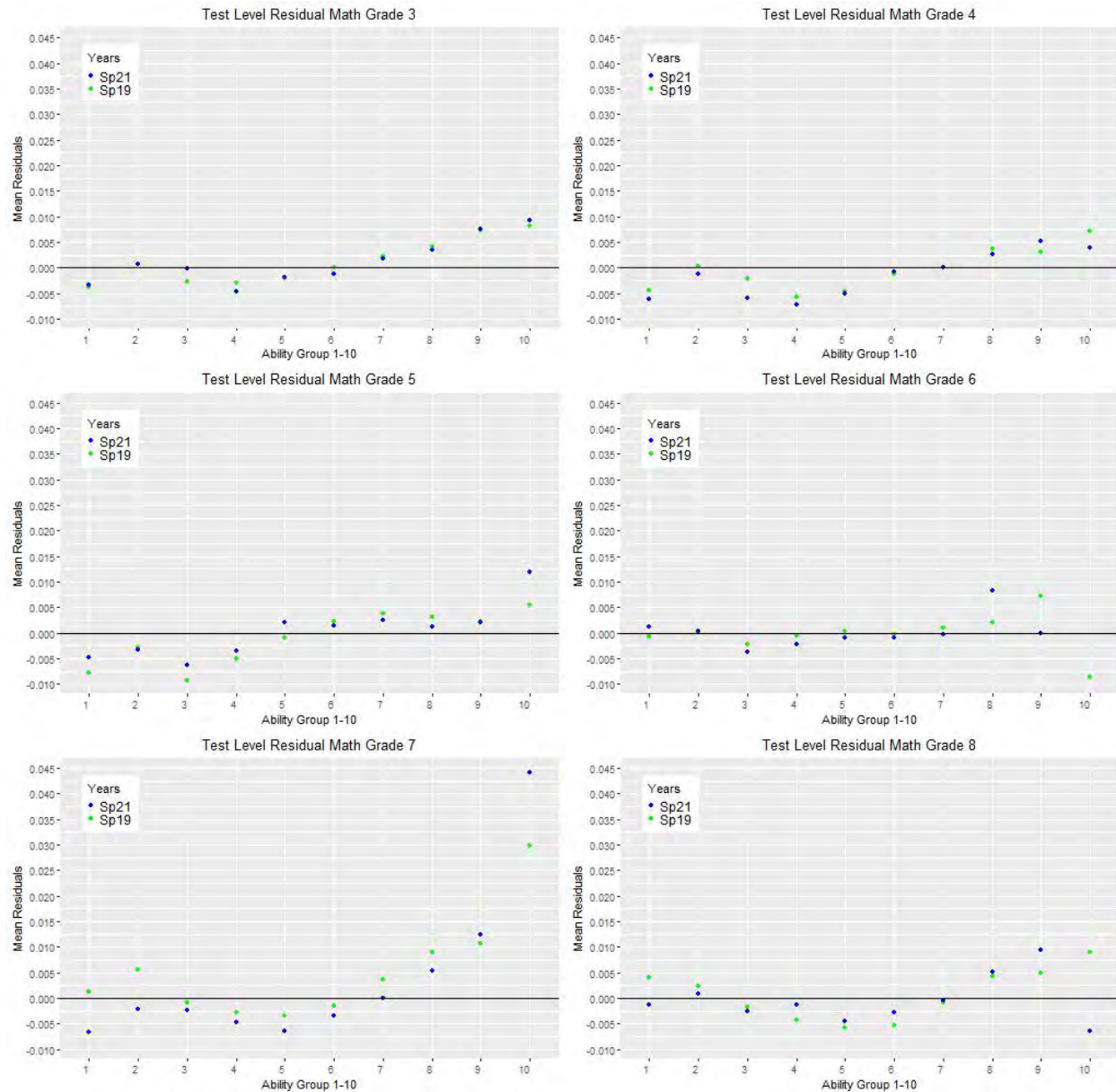


Figure 2: Math Test Level Residual Comparison

6. Correlations

To ensure that the predictive validity of the WVGSA was not impacted by the conditions brought about by the pandemic, correlational studies were conducted between the 2021 and 2019 data both at the individual student level and at aggregate levels to examine the predictive strength of the scores from 2019 to the scores from 2021. Changes in discriminant validity was also monitored by correlations between ELA and math for 2018, 2019, and 2021. (See Technical Manual, Vol. 4 for a more complete discussion). If any of these correlational relationships had been shown to be lower than expected, this would provide potential areas for future research into the extent of the impact of the pandemic on student learning. It should be noted that grades 5 – 8 were the only ones to be included in these analyses as students in 4th grade in spring 2020 were not tested that year due to the pandemic.

Table 16: West Virginia Math Scale Score Correlations SP18-SP19

SP18_Grade	SP19_Grade	N	correlation
3	4	18330	0.83
4	5	18699	0.83
5	6	18670	0.81
6	7	17900	0.83
7	8	18007	0.82

Table 17: West Virginia Math Scale Score Correlations SP19-SP21

SP19_Grade	SP21_Grade	N	correlation
3	5	15698	0.79
4	6	15732	0.79
5	7	16146	0.79
6	8	16147	0.77

Table 18: West Virginia ELA Scale Score Correlations SP18-SP19

SP18_Grade	SP19_Grade	N	correlation
3	4	18327	0.78
4	5	18693	0.80
5	6	18662	0.80
6	7	17884	0.82
7	8	17984	0.82

Table 19: West Virginia ELA Scale Score Correlations SP19-SP21

SP19_Grade	SP21_Grade	N	correlation
3	5	15639	0.76
4	6	15637	0.77
5	7	16052	0.79
6	8	16043	0.79

Table 20: West Virginia Correlations Between ELA and Math Scale Scores

Grade	SP18_N	SP18_Correlation	SP19_N	SP19_Correlation	SP21_N	SP21_Correlation
3	19336	0.75	18545	0.75	16292	0.74
4	19741	0.74	19199	0.76	16581	0.74
5	19870	0.74	19533	0.76	16734	0.75
6	19070	0.78	19564	0.76	16788	0.75
7	19267	0.75	18812	0.74	17149	0.74
8	19356	0.74	19013	0.76	17145	0.73

Results

First, it is important to note that although some differences in performance between the spring of 2021 and previous years were detected, variation across testing occasions is expected, and it is therefore difficult to attribute these changes solely to the circumstances brought on by the pandemic. It was not possible to separate all possible sources of variation with the available data. Thus, the following results might best be used to gain insight into areas where additional investigations may be warranted to better understand how and why these changes may have occurred.

Proficiency Levels. Across grades in both ELA and math, the percentage of students that scored above the proficiency cut score was shown to decline, with the biggest impact being observed in math proficiency levels. In ELA, proficiency levels dropped between 1% for 8th grade students and 11% for 4th grade students. Math students of all grades saw at least a 10% decline in proficiency levels, with the greatest decline of 13% found for 3rd, 4th, and 6th grade students.

Scale Scores. Like proficiency levels, mean scale scores were also shown to decline across grades for both ELA and math. In ELA, mean scale scores dropped between 2 points for 8th grade students to 11 points for both 3rd and 4th grade students. In math, 3rd grade students (11 points) showed the least amount of impact while the greatest decline (21 points) was shown for 8th grade students. When spring 2021 scale scores were compared to the mean scores of both spring 2018 and spring 2019, the changes in mean differences were similar, ranging from a 1 – 12 point drop in ELA scale scores and a 10 – 20 point drop in math scale scores.

Reliability. Minimal changes in reliability were observed across all grades for both ELA and math. In math, the 7th grade test showed no change in test reliability and the largest decrease ($\alpha = -0.03$) was found for the 5th grade math test. In ELA, 3rd, 6th, 7th, and 8th grade tests showed no change in reliability. The 4th ($\alpha = -0.01$) and 5th ($\alpha = -0.03$) grade assessment reliability coefficients decreased only slightly.

Item Statistics. In ELA, the percent of items across content standards that demonstrated significant increases in item difficulty ranged from 5% to 100%. Grade 8 showed the lowest number of content standards containing flagged items (7) and Grade 4 showed the highest number (14). In math, the percent of content standards with significant increases in item difficulty ranged from 5% to 100%. Grades 7 and 8 showed the lowest number of content standards containing flagged items (17) and Grade 5 showed the highest number (25).

Residual Analyses. In general, the residual patterns for spring 2021 follow the patterns found in spring 2019. In most of the cases, the mean residuals are in the same direction (plus or minus) or are close to zero (the expected average). Considering the variability observed in the past years, the changes between spring 2019 and spring 2021 in cohort groups do not appear to bring concerns related to test score validity/stability of the WVGSA during the pandemic. These non-significant changes in residuals accompanied by the observed declines in test scores (section 1 and section 2) indicate the effectiveness of the CAI CAT item selection algorithm in tailoring item difficulty level to students whose academic performance might have been interrupted by the pandemic.

Correlations. Correlations between the scale scores between spring 2019 and spring 2021 (ranging from $\rho = 0.76$ to $\rho = 0.79$ for ELA, and from $\rho = 0.77$ to $\rho = 0.79$ in math), showed little change compared to the

correlations between scale scores between spring 2018 and spring 2019 (ranging from $\rho = 0.78$ to $\rho = 0.82$ for ELA, and from $\rho = 0.81$ to $\rho = 0.83$ in math). Correlations between math and ELA also saw very little change since the spring 2018 and spring 2019.

References

- Boyer, M., Dadey, N., Keng, L. (2020). Statewide summative assessment in spring 2021: A workbook to support planning and decision-making. Dover, NH: National Center for the Improvement of Educational Assessment.
- Council of Chief State School Officers (2020a). Restart and Recovery: Assessments in Spring 2021. Washington, DC: Author.
- Council of Chief State School Officers (2020b). Restart and Recovery: Accountability interrupted: Guidance for collecting, evaluating, and reporting data in 2020-2021. Washington, DC: Author.
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), Educational Measurement (pp. 65 – 110). Westport, CT: Praeger Publishers.

Equating Test Scores across Administrations in the Context of the Covid-19 Pandemic: An Investigation into the Impact of Gaps in Student Learning on Field Test Item Calibration

Abstract

Since the onset of the pandemic, concerns have been raised regarding how instructional changes in K-12 educational content might impact field-testing and the validity of equating results for large-scale assessment programs. The current study investigated the potential impact of any gaps in student learning on field-test item calibration and post-equating results for the Independent College and Career Readiness (ICCR) item bank. Item fit analysis was conducted on the operational items from the ICCR item bank and flagged if RMSEA was greater than 0.05. Academic content standards for ELA and math were then flagged if they contained 50% or more misfitting items. Item parameters for field-test items were estimated in two separate calibrations from the spring 2021 test administration using data from the four states that utilize the ICCR item bank. First, field test items were calibrating using the operational items with fixed parameters as anchor items. Second, all operational and field-test items from flagged content standards were removed, and the retained field-test items were recalibrated using the remaining operational items as anchors. Item parameters and the standard errors of estimation (SEE) from the two calibrations were compared, finding that they were highly correlated across all scoring models, and that no differences between the parameter SEEs were greater than the significance criterion of 1. Additionally, the removal of ELA items impacted calibration convergence for grades 3, 4, 8, and 9. With this, the decision was made to retain the field-test items from the spring 2021 administration and from the original calibration.

Introduction

Over the last two years, students and teachers have experienced unprecedented educational challenges due to school closures and disruptions in student learning due to the pandemic. There was a concern that there have been changes in the opportunity to learn (OTL) the content being tested due to a lack of instructional coverage of content, differences in test modalities (remote versus onsite testing), and barriers to learning such as a lack of internet access and varying levels of instructional and emotional support at home (Boyer, Dadey, & Keng, 2020). Serious changes in OTL can impact operational assessment procedures, particularly where field testing and equating are concerned (Keng, Boyer, & Marion, 2020).

Equating procedures are used to ensure the equivalence of test score interpretations across test administrations, allowing educators to monitor student achievement of academic content standards across test administrations and cohorts. Both pre-equating and post-equating methods are used in statewide assessment programs. The principal advantage of pre-equating methods is that they can be applied before the assessment takes place, allowing for immediate scoring and reporting of test results. In addition, they allow the possibility of computer adaptive item selection. However, pre-equating methods rely on the assumption that the item parameters are invariant across measurement occasions. This assumption may be violated, for example, when items are calibrated several years prior to their current use (item parameter drift over time).

Post-equating methods are applied after the operational test administration. They rely on a set of anchor items, a representative set of items that are in common between the current test and a previous test (or item bank). After the operational test administration, all items are (re)calibrated, and the anchor items are used to identify a linear transformation of the new item parameters so that they are expressed on the

existing scale. Whereas pre-equating methods rely on the assumption that the psychometric properties of items did not change from one measurement occasion to another, it is sufficient for post-equating methods that the assumption of item parameter invariance holds for the common items. The item parameters of all other items are always based on the most current administration.

In practice, a mixed approach is often used. For example, even in the context of an adaptive assessment system, the pre-equated item parameters are used for item selection and scoring of all operational test items, but items that show item drift can be recalibrated and post-equated to the base scale using the remaining items as common items.

As noted above, adaptive test administrations require the use of pre-equated item parameters to enable adaptive item selection. More generally, one of the most consequential features of online test administration, coupled with automated scoring of test items, is the ability to score and report assessment results immediately or within a short reporting frame. In fact, students and educators participating in online assessments expect to view assessment results immediately or with very little delay. For these reasons, the slight advantages accrued by post-equating may be outweighed by delays in score reporting. Thus, for purposes of both adaptive test administration and immediate or more rapid reporting of assessment results, Cambium Assessment Inc. (CAI) typically employs pre-equated item parameters for item selection and for scoring and reporting of assessment results.

The use of pre-equated item parameters for item selection and the scoring and reporting of test results has additional benefits in the context of the Covid-19 pandemic. The typical argument for post-equating is that post-equating accounts for all local effects of a test administration, including for example, item position, current events, and other factors that might make the item easier or more difficult. In this way, the post-equated item parameters better reflect the difficulty of the items in the current assessment context. However, due to very unusual and disruptive effects on education caused by the pandemic, the model parameters would not simply be refined to reflect small differences in the test form or administration but rather the larger impacts on instruction and learning due to the pandemic, including lost instructional time and remote learning. To accurately identify the impact of the pandemic on student achievement, we argue that the most accurate and reliable item parameters for scoring students' assessments are those estimated prior to the pandemic.

The CAI field test engine deployed for adaptive test administrations randomly selects items from the field test pool for administration within the summative test, and randomly positions items within a range of allowable item positions (typically not the first or last five items in a test administration). Thus, all field test items are administered to a random sample of students statewide, and each item is administered across a wide range of possible item positions, averaging out any effects of item position in item parameter estimates. Since pre-equated item parameters in these assessment systems are based on responses to field test items that are embedded in the operational test administration, item parameter estimates are robust and capture all factors influencing student responses in operational testing conditions.

Usually field test item parameters were estimated by anchoring all of the operational item parameters during calibration of the item parameters which put all field test item parameters on the same scale as the operational item parameters. In this report, we looked at the impacts of removing some of the anchor items flagged by the residual analysis.

Monitoring Item Drift

Although all operational scoring of test items is based on pre-equated item parameters, it is nevertheless important to evaluate item parameters for drift over time. Item parameter drift (IPD) can occur when items become compromised due to item over exposure, or as a result of gradual shifts in curriculum (Guo, Robin, & Dorans, 2017). In large scale tests, the pre-equating methods such as those used for the Independent College and Career Readiness (ICCR) item bank rely on the assumption of stable item parameters across testing administrations, and any variable that poses a threat to the stability of the anchor item parameters also poses a threat to the validity of equating results (Guo, Zheng, & Chang, 2015). Generally, if item drift occurs at all, it usually does so over a long period of time. However, Guo et al. (2017) also note that

IPD may also develop quickly under some circumstances, and cite social events in addition to overexposure and curriculum changes as possible reasons for IPD. When IPD is detected under typical testing conditions, items can be re-field-tested to update the item parameters, or the items can be targeted for transition to an interim assessment or practice test item pool.

The effects of disruption of in-school instruction due to the Covid-19 pandemic on student achievement are not known, but it seems likely that lost instruction time and remote learning will result in a general decline in academic achievement rather than a differential pattern of decline across specific content standards. A general decline in student achievement would not be expected to result in a decreased fit of items to the item response theory (IRT) model. Item misfit, however, would be expected to occur in the context of more targeted shifts in instruction. For example, if some content standards are no longer a target of instruction statewide due to decreased instructional time or limitations of remote instruction, then otherwise high achieving students who would typically have performed well on that content might no longer perform as expected due to lack of instruction. Such systematic shifts in instruction across the population would be expected to decrease item fit to the model parameters. We note that while loss of instructional time may result in content specific impacts on instruction, the content standards impacted could vary across classrooms, schools, and districts. Although such local impacts on instruction would be expected to impact student learning, we would expect to see evidence of variable instructional impacts on person fit indices rather than item fit.

Field Testing

The addition of field test items each year is a critical component to the maintenance and replenishment of the ICCR item bank, therefore a limited number of items were field tested during the Spring 2021 test administration. Under the present conditions, it has been suggested that the level of concern over field testing should be dictated by the role of field test items in the test development process (Keng, Boyer, & Marion, 2020). In a post-equating model, such as that used for the ICCR item bank, field test item quality is less consequential than in a pre-equating model where field test items are used to establish the scoring tables for operational items.

Since field test items are calibrated by anchoring operational test items to their bank values, field test item parameter estimates can be equated to the bank scale. However, should our analyses of item fit identify greater than expected rates for specific content standards, field test items measuring those standards would be excluded from the item calibrations, and field tested again in a future test administration. Additionally, operational items measuring those standards would also be excluded from the set of anchor items to ensure that ability estimates for students administered items measuring those standards are not biased. We acknowledge that this approach assumed limited or no evidence of systematic item parameter drift. If evidence of item misfit is found across a range of content standards, then items would need to be field tested again in a subsequent test administration.

Method

Field test data from all four states using the ICCR item bank was included in the present study. In math, across all states, approximately 25 standalone items and 20 cluster items were administered per grade. In ELA, approximately 60 standalone 7 cluster items per grade were administered. In New Hampshire, only cluster items were field tested, West Virginia only tested standalone items, and Wyoming and North Dakota field tested both item types. Only the standalone items were included in the present study.

In the first calibration, all field test item parameters were estimated using the operational items as anchors with fixed parameters. Using the response data from the four states that utilize the ICCR item bank, RMSEA fit statistics were calculated using the following methods for all operational items used in the spring 2021 test administration. Items were flagged as misfitting the scoring model if the RMSEA value was greater than 0.05. Any items with less than 100 responses were removed from the remaining analyses.

ELA and math content levels were evaluated according to the percentage of misfitting items. Content standards found to contain 50% or more misfitting items were flagged for removal from the second calibration. All field test and anchor items found within the flagged content standards were then removed and the remaining items were recalibrated.

Item Displacement

For the Rasch based IRT model, IPD can be measured using item displacement indices. For a general IRT model, we can define the displacement based on the bank parameters, student ability estimates, and the item scores.

If there are n students who have taken the item, let z_i represent the i th student's score on the item, θ_i be the estimated ability, and the item parameters be a , b , and c if the item is dichotomous item, and a and $b_1, \dots, b_M$ for maximum possible item score of M starting from item score of 0 for polytomous item. Let D be the IRT scale factor, which is, in general, 1 for 1PL model, and 1.7 for the 2PL and 3PL models.

If the item is a dichotomous item, the expected score for student i is estimated as

$$E(z_i) = c + (1 - c) \frac{\exp(Da(\hat{\theta}_i - b))}{1 + \exp(Da_j(\theta_i - b))}$$

If the item is a polytomous item, the expected score is estimated as

$$E(z_i) = \sum_{l=1}^M \frac{l \text{Exp}(Da \sum_{k=1}^l (\theta_i - b_k))}{1 + \sum_{m=1}^M \text{Exp}(Da \sum_{k=1}^m (\theta_i - b_k))}$$

The item score variance is estimated by

$$\text{Var}(z_i) = E(z_i)(1 - E(z_i))$$

if the item is a dichotomous item, and

$$\text{Var}(z_i) = \sum_{l=1}^M \frac{l^2 \text{Exp}(Da \sum_{k=1}^l (\theta_i - b_k))}{1 + \sum_{m=1}^M \text{Exp}(Da \sum_{k=1}^m (\theta_i - b_k))} - \left(\sum_{l=1}^M \frac{l \text{Exp}(Da \sum_{k=1}^l (\theta_i - b_k))}{1 + \sum_{m=1}^M \text{Exp}(Da \sum_{k=1}^m (\theta_i - b_k))} \right)^2$$

if the item is a polytomous item.

The residual between the observed and expected scores for student i is defined as

$$\delta_i = E(z_i) - z_i$$

The average residual is

$$\bar{\delta} = \frac{\sum_{i=1}^n \delta_i}{n}$$

and the average variance is

$$\bar{v} = \frac{\sum_{i=1}^n Var(z_i)}{n},$$

then item displacement is estimated by

$$d = \frac{\bar{\delta}}{\bar{v}}$$

Item can be flagged using $d > .5$ or $.3$ when Rasch model is used.

Item Fit

The general approach to evaluating goodness-of-fit involves the comparison between observed and model-predicted frequencies for various ability (theta) subgroups using chi-square fit statistics. However, the ability is a latent variable and theta values cannot be directly observed. Two different ways to deal with this problem are (1) using summed raw score and (2) using pseudo-observed score distribution. The summed raw score cannot be used for CAT data where students get different sets of items. Therefore, the item fit statistics we propose employ the observed theta distribution.

The steps to calculate the chi-square fit statistics include:

- 1) Grouping the n students taking the j th item into K (for example, $K=10$) groups with similar n -counts using their estimated ability. Let the min and max theta for the k th group be θ_{1k} and θ_{2k} , we can form K intervals based on

$$\theta_k = \frac{\theta_{1k+1} + \theta_{2k}}{2}$$

for

$$k = 1, 2, \dots, K - 1.$$

Let the MLE ability be θ_i and SEM be $sem(\theta_i)$ for the i th person. The probability of the i th person's ability in the k th interval is estimated by:

$$\text{for } k = 1, p_i(k) = \Phi\left(\frac{\theta_1 - \hat{\theta}_i}{sem(\hat{\theta}_i)}\right),$$

$$\text{for } 1 < k < K, p_i(k) = \Phi\left(\frac{\theta_k - \hat{\theta}_i}{sem(\hat{\theta}_i)}\right) - \Phi\left(\frac{\theta_{k-1} - \hat{\theta}_i}{sem(\hat{\theta}_i)}\right),$$

and $k = K$, $p_i(k) = 1 - \Phi\left(\frac{\theta_{K-1} - \hat{\theta}_i}{sem(\hat{\theta}_i)}\right)$,

where ϕ is the conditional density function (CDF) of the standard normal distribution.

- 2) In each ability group, compute the pseudo-count of the observed students and expected students using IRT models.

If the item is a MC item, then the probability of the item score for the i th student is estimated by

$$p_{im} = 1 - c - (1 - c) / \left(1 + \text{Exp}\left(-Da(\hat{\theta}_i - b)\right)\right)$$

for $m = 0$

$$p_{im} = c + (1 - c) / \left(1 + \text{Exp}\left(-Da(\hat{\theta}_i - b)\right)\right)$$

for $m = 1$. If the item is a polytomous item, then for $m=0$

$$p_{im} = \frac{1}{1 + \sum_{l=1}^M \text{Exp}\left(\sum_{k=1}^l Da(\hat{\theta}_i - b_k)\right)}$$

and $m > 0$

$$p_{im} = \frac{\text{Exp}\left(\sum_{k=1}^m Da(\hat{\theta}_i - b_k)\right)}{1 + \sum_{l=1}^M \text{Exp}\left(\sum_{k=1}^l Da(\hat{\theta}_i - b_k)\right)}$$

The pseudo-count of the observed students at item score of m in the k th interval is estimated by

$$O_{mk} = \sum_{i=1}^n p_i(k) \times 1_{(r_i=m)}$$

where $1_{(r_i=m)} = \begin{cases} 1 & \text{if } r_i = m \\ 0 & \text{if } r_i \neq m \end{cases}$

The pseudo-count of the expected students at item score of m at θ_k is estimated by

$$E_{mk} = \sum_{i=1}^n p_{im} p_i(k)$$

3) Comparing observed and expected distributions and exam the residual for each of items

The χ^2 or Q_1 is estimated by

$$\chi^2 = \sum_{k=1}^K \sum_{m=0}^M \frac{(O_{mk} - E_{mk})^2}{E_{mk}}$$

and G^2 is estimated by

$$G^2 = 2 \sum_{k=1}^K \sum_{m=0}^M O_{mk} \log \left(\frac{O_{mk}}{E_{mk}} \right)$$

χ^2 or G^2 are sensitive to sample size. To account for sample size, χ^2 and G^2 are transformed to root mean square error of approximation (RMSEA) by

$$\sqrt{\frac{\chi^2 - df}{df(n-1)}}$$

and flag the item if RMSEA > .05, where df = KM - n_p , and n_p is the number of parameters.

Review by Content Experts

The flagged operational items and content standards were sent for review by CAI content experts to determine if any apparent causes for the item misfit could be identified based on their expertise on the subject.

Standard Error of the Estimate

Standard error of the estimate (SEE) is an index that provides an indication of how well an item parameter has been estimated (i.e. difficulty parameter (b), discrimination parameter (a)). Much like a standard deviation, this value serves as an index of the estimator's variability, with larger values indicating a higher degree of error in the estimate of the parameter (de Ayala, 2009).

For both calibrations, items parameters were inspected and labeled as not to be used in future test administrations if the b parameter was greater than 6 or if the a parameter was greater than 5 or less than 0.01. The SEEs of the item parameters obtained from this second calibration were compared with the SEEs from the original calibration containing all items. Any differences in SEEs greater or equal to the absolute value of 1 would be considered a significant difference in the estimation of the item parameter.

Results

In ELA, the percentage of operational items flagged as misfitting ranged from 27.50% of 10th grade items to 33.58% of 7th grade items (Table 1). Similar proportions were observed in the math operational items, with proportions of misfit ranging from 24.8% of 4th grade operational items to 39.4% of 8th grade items (Table 2). Field-test items were considered flagged if they fell into a content area that was found to contain

at least 50% misfitting items. The highest percentage (10.11%) of flagged ELA field-test items was observed in 3rd grade items while none of the 6th grade field-test items fell into a flagged content area. The highest percentage (36.36%) of math field-test items was found for 9th grade items, although this grade had the fewest field-test items ($n = 11$) across states with only 4 flagged items. This was followed closely by 10th grade field-test items where 33.33% were flagged as misfitting. None of the 7th grade math field-test items fell into a flagged content area. A total of 39 content levels for ELA and 71 content levels for math were flagged as containing 50% or more misfitting items (Tables 3 and 4).

Table 1: Flagged ELA Operational and Standalone Field Test Items

Grade	N OP Items	N OP Items Flagged	Percent OP Items Flagged	N FT Items	N FT Items Flagged	Percent FT Items Flagged
3	443	137	30.93%	89	9	10.11%
4	463	133	28.73%	72	1	1.39%
5	451	120	26.61%	79	3	3.8%
6	538	150	27.88%	51	0	0%
7	533	179	33.58%	63	3	4.76%
8	495	139	28.08%	84	1	1.19%
9	260	70	26.92%	40	4	10%
10	280	77	27.5%	46	1	2.17%

Table 2: Flagged Math Operational and Standalone Field Test Items

Grade	N OP Items	N OP Items Flagged	Percent OP Items Flagged	N FT Items	N FT Items Flagged	Percent FT Items Flagged
3	601	158	26.29%	41	3	7.32%
4	637	158	24.8%	37	3	8.11%
5	530	173	32.64%	38	8	21.05%
6	641	196	30.58%	44	2	4.55%
7	454	154	33.92%	33	0	0%
8	500	197	39.4%	40	10	25%
9	358	117	32.68%	11	4	36.36%
10	353	113	32.01%	36	12	33.33%

Table 3: Flagged ELA Content Standards

ICCR Content Standard	N Items	N Items Flagged	Percent Flagged
AIR-ELA-v2:L.3.1e	6	4	67%
AIR-ELA-v2:L.3.1g	6	5	83%
AIR-ELA-v2:L.3.1h	6	3	50%
AIR-ELA-v2:L.3.2c	2	1	50%
AIR-ELA-v2:L.3.2e	6	3	50%
AIR-ELA-v2:L.3.4b	3	3	100%
AIR-ELA-v2:L.3.5a	17	9	53%
AIR-ELA-v2:L.4.1f	5	3	60%
AIR-ELA-v2:L.4.1g	4	2	50%
AIR-ELA-v2:L.4.2a	9	7	78%
AIR-ELA-v2:L.4.5c	3	2	67%

ICCR Content Standard	N Items	N Items Flagged	Percent Flagged
AIR-ELA-v2:L.5.1c	3	3	100%
AIR-ELA-v2:L.5.1e	8	4	50%
AIR-ELA-v2:L.5.2a	4	2	50%
AIR-ELA-v2:L.5.5c	1	1	100%
AIR-ELA-v2:L.6.1b	3	2	67%
AIR-ELA-v2:L.6.1e	2	1	50%
AIR-ELA-v2:L.7.4b	4	2	50%
AIR-ELA-v2:L.7.5b	1	1	100%
AIR-ELA-v2:L.7.5c	4	2	50%
AIR-ELA-v2:L.8.5c	2	1	50%
AIR-ELA-v2:L.9-10.5a	8	4	50%
AIR-ELA-v2:RI.3.4	27	14	52%
AIR-ELA-v2:RI.7.7	2	1	50%
AIR-ELA-v2:SL.6.2	7	4	57%
AIR-ELA-v2:SL.6.3	2	1	50%
AIR-ELA-v2:SL.7.3	1	1	100%
AIR-ELA-v2:SL.8.3	3	2	67%
AIR-ELA-v2:SL.9-10.3	1	1	100%
AIR-ELA-v2:W.3.1a	6	6	100%
AIR-ELA-v2:W.3.2a	9	7	78%
AIR-ELA-v2:W.4.1a	6	3	50%
AIR-ELA-v2:W.4.2a	6	3	50%
AIR-ELA-v2:W.5.1a	9	5	56%
AIR-ELA-v2:W.7.2a	6	5	83%
AIR-ELA-v2:W.8.1a	6	3	50%
AIR-ELA-v2:W.8.2a	6	3	50%
AIR-ELA-v2:W.9-10.1a	6	4	67%
AIR-ELA-v2:W.9-10.2a	6	4	67%

Table 4: Flagged Math Content Standards

ICCR Content Standard	N Items	N Items Flagged	Percent Flagged
AIR-MA-v2:3.MD.6	8	4	50%
AIR-MA-v2:3.MD.7c	1	1	100%
AIR-MA-v2:3.OA.6	20	17	85%
AIR-MA-v2:4.MD.1	17	13	76%
AIR-MA-v2:4.NF.4a	5	3	60%
AIR-MA-v2:4.NF.4c	13	7	54%
AIR-MA-v2:5.MD.3b	4	2	50%
AIR-MA-v2:5.MD.5a	4	2	50%
AIR-MA-v2:5.MD.5b	10	6	60%
AIR-MA-v2:5.MD.5c	4	2	50%
AIR-MA-v2:5.NBT.1	16	8	50%
AIR-MA-v2:5.NF.5b	6	3	50%
AIR-MA-v2:5.OA.3	19	10	53%
AIR-MA-v2:6.NS.7b	4	4	100%
AIR-MA-v2:6.RP.3b	20	14	70%
AIR-MA-v2:6.SP.5a	3	2	67%
AIR-MA-v2:7.G.3	14	10	71%
AIR-MA-v2:7.G.4	17	13	76%
AIR-MA-v2:7.NS.1d	11	6	55%
AIR-MA-v2:7.NS.2a	4	3	75%
AIR-MA-v2:7.NS.2d	8	7	88%
AIR-MA-v2:7.RP.1	30	15	50%
AIR-MA-v2:7.SP.6	11	6	55%
AIR-MA-v2:7.SP.7b	5	4	80%
AIR-MA-v2:7.SP.8a	3	2	67%
AIR-MA-v2:8.EE.4	11	7	64%
AIR-MA-v2:8.EE.7b	5	3	60%
AIR-MA-v2:8.F.1	20	11	55%
AIR-MA-v2:8.F.5	22	13	59%
AIR-MA-v2:8.G.1a	9	5	56%
AIR-MA-v2:8.G.1b	3	3	100%
AIR-MA-v2:8.G.3	11	7	64%
AIR-MA-v2:8.NS.1	20	16	80%
AIR-MA-v2:8.NS.2	29	18	62%
AIR-MA-v2:8.SP.2	18	11	61%
AIR-MA-v2:A-CED.1 Exp	1	1	100%
AIR-MA-v2:A-CED.1 Lin	9	7	78%
AIR-MA-v2:A-CED.3 Lin	2	1	50%
AIR-MA-v2:A-CED.3 Quad	1	1	100%
AIR-MA-v2:A-CED.4 Lin	1	1	100%
AIR-MA-v2:A-REI.11 Quad	2	1	50%
AIR-MA-v2:A-SSE.1a EInt	3	2	67%
AIR-MA-v2:A-SSE.1a Rat	2	1	50%
AIR-MA-v2:A-SSE.3a Quad	3	2	67%
AIR-MA-v2:F-BF.1a	1	1	100%
AIR-MA-v2:F-BF.1a Exp	2	2	100%
AIR-MA-v2:F-BF.3 Quad	2	1	50%
AIR-MA-v2:F-IF.1	5	4	80%
AIR-MA-v2:F-IF.1 Lin	2	1	50%
AIR-MA-v2:F-IF.2 EInt	4	3	75%

ICCR Content Standard	N Items	N Items Flagged	Percent Flagged
AIR-MA-v2:F-IF.3 EInt	1	1	100%
AIR-MA-v2:F-IF.4 Quad	4	3	75%
AIR-MA-v2:F-IF.5 Lin	2	1	50%
AIR-MA-v2:F-IF.5 Rad	2	1	50%
AIR-MA-v2:F-IF.6 Quad	4	2	50%
AIR-MA-v2:F-IF.6 Rad	1	1	100%
AIR-MA-v2:F-IF.9 Lin	1	1	100%
AIR-MA-v2:F-IF.9 PW	1	1	100%
AIR-MA-v2:F-LE.2 Lin	3	2	67%
AIR-MA-v2:F-LE.3 Quad	5	3	60%
AIR-MA-v2:F-LE.5 Lin	1	1	100%
AIR-MA-v2:G-CO.1	6	3	50%
AIR-MA-v2:G-CO.10	6	3	50%
AIR-MA-v2:G-CO.11	6	3	50%
AIR-MA-v2:G-GMD.1	6	3	50%
AIR-MA-v2:G-GMD.3	7	5	71%
AIR-MA-v2:G-MG.2	12	6	50%
AIR-MA-v2:G-SRT.6	9	8	89%
AIR-MA-v2:S-CP.2	3	3	100%
AIR-MA-v2:S-ID.3	5	3	60%
AIR-MA-v2:S-ID.6a	2	2	100%

Content Expert Reviews

It should first be noted that since the content standards and the items that measure them have not changed, the content experts involved in this review did not expect that any changes in alignment to the standards would have occurred. Therefore, the items and corresponding standards were reviewed for any observed patterns that might be suggestive of changes in learning. Their thoughts on reasons for such changes were speculative as these changes are likely to vary by state, and any data that might account for these observations is currently unavailable.

The content expert who reviewed the ELA flagged items and standards observed that 90% of the flagged items were found in the areas of grammar, vocabulary, and writing. As this type of instruction often takes place as a collaborative one-to-one interaction with the instructor, it was speculated that the pandemic related learning conditions may have compromised instruction and practice in these areas. It was found notable that a high number of editing task items, 14 out of 27 items from standard RI.3.4 (described as the most surprising observation), several items that required listening, and all writing items had been flagged.

In contrast, the content expert who reviewed the flagged items and standards in mathematics was unable to observe any clear trends. It was initially expected that the standards representing statistics, geometry, or abstract concepts might exhibit proportionally more misfit since these are topics often covered later in the school year under normal learning conditions. However, it was observed that most math domains across grades have at least some degree of misfit, and although statistics and geometry standards exhibited slightly more misfit than other standards, it was not enough to constitute a clear trend.

Comparisons of the Calibration Results

The removal of the items based on the RMSEA flagging criteria was found to have a minimal impact on both math and ELA a and b parameter SEEs for all retained field test items. Differences in the a and b item parameters between calibrations for all models, and the c parameter for the 3PL model, were also investigated and found to be minimal. The differences in the ELA 2PL item parameters ranged from -0.46 to 0.25, the ELA 3PL differences ranged from -1.06 to 0.87, and the GPC from -0.15 to 0.08. The math 2PL

item parameter differences ranged from -0.9 to 0.22, the 3PL parameter differences from -0.54 to 0.56, and -0.04 to 0.14 for the GPC (Tables 3 and 4). The item parameters from the separate calibrations were also found to be highly correlated (Figures 1 - 14), with the smallest correlation being found between the ELA 3PL c parameters ($r = 0.95$).

The removal of items prior to recalibration was found to impact calibration convergence for ELA in grades 3, 4, 8, and 9, but did not impact convergence for any grades in math recalibration. A common observation in the affected grade levels was that all operational writing items had been dropped from the anchor set. Interestingly, all writing items were also dropped from the grade 10 anchor set but were found to have no impact on calibration convergence.

Table 5: ELA Item Parameter Comparison

Parameters	Original.calibration	Recalibration
A parameter		
Mean	1.58	1.62
SD	0.61	0.68
SEE	0.2	0.2
SEE Range	0.06 - 1.13	0.05 - 1.46
B parameter		
Mean	0.29	0.3
SD	1.03	0.98
SEE	0.15	0.15
SEE Range	0.03 - 1.97	0.03 - 1.92
C parameter		
Mean	0.2	0.19
SD	0.08	0.08
Mean SEE	6.37	7.27
SEE Range	3.00 - 18.00	3.00 - 18.00

Table 6: Math Item Parameter Comparison

Parameters	Original.calibration	Recalibration
A parameter		
Mean	1.65	1.62
SD	0.76	0.76
SEE	0.2	0.19
SEE Range	0.05 - 1.17	0.05 - 0.95
B parameter		
Mean	0.34	0.33
SD	1.18	1.22
SEE	0.12	0.12
SEE Range	0.03 - 0.47	0.03 - 0.52
C parameter		
Mean	0.19	0.19
SD	0.08	0.09
Mean SEE	5.11	5.15
SEE Range	3.00 - 12.00	3.00 - 12.00

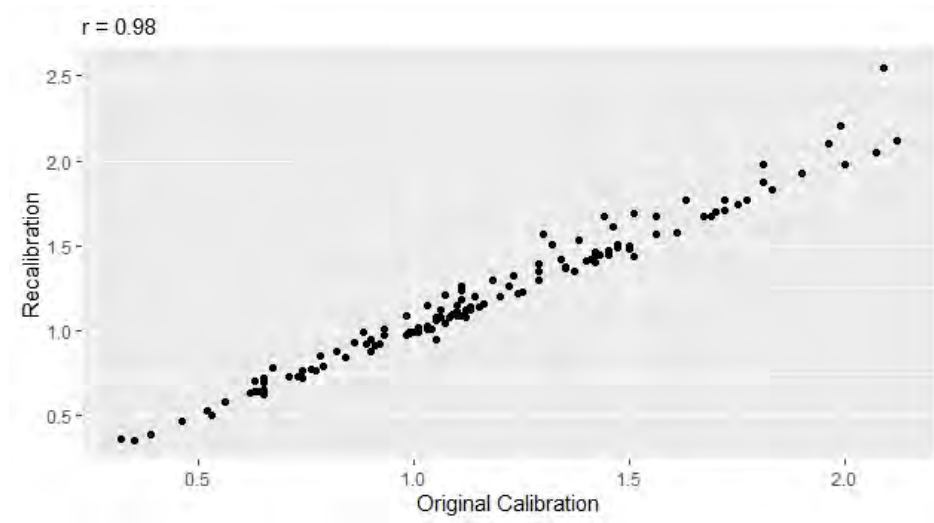


Figure 1: Correlation Between ELA 2PL A Parameters

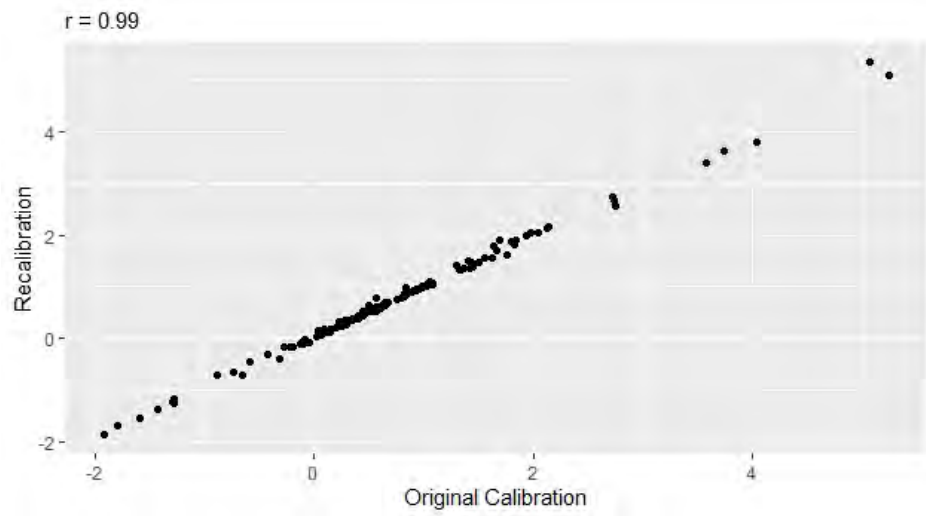


Figure 2: Correlation Between ELA 2PL B Parameters

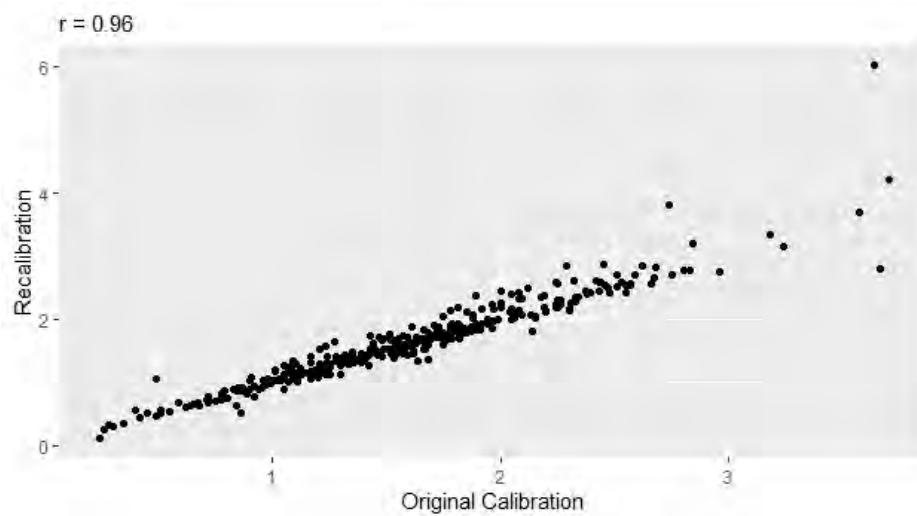


Figure 3: Correlation Between ELA 3PL A Parameters

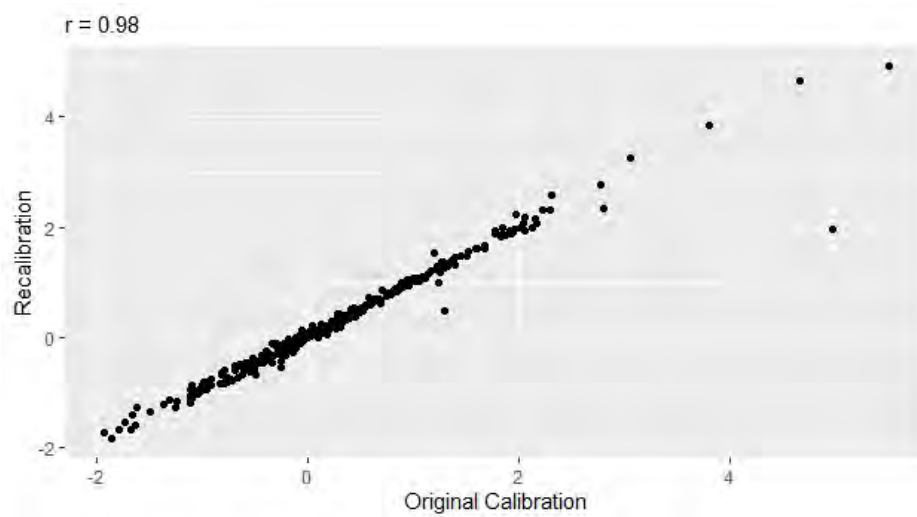


Figure 4: Correlation Between ELA 3PL B Parameters

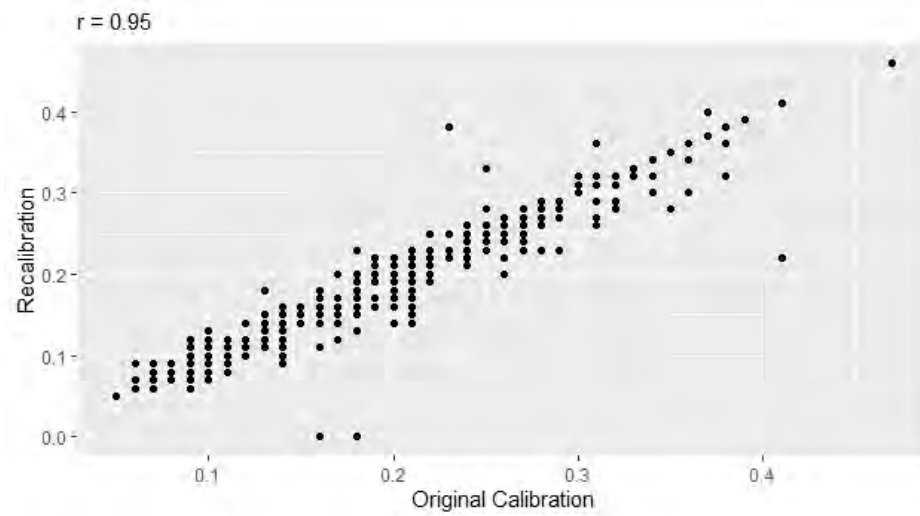


Figure 5: Correlation Between ELA 3PL C Parameters

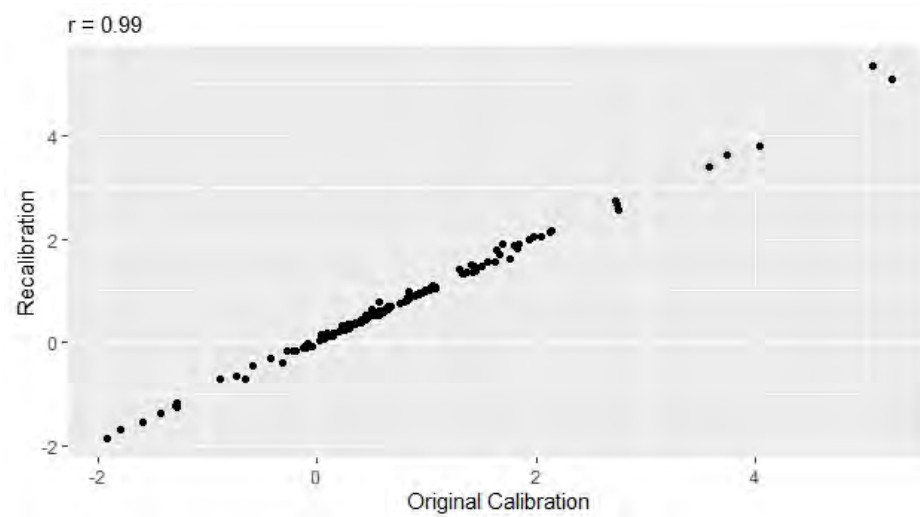


Figure 6: Correlation Between ELA GPC A Parameters

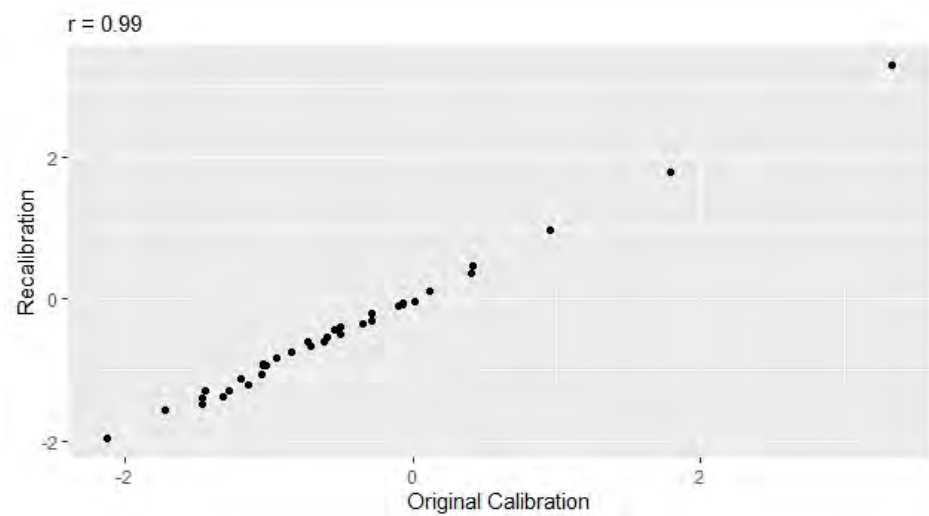


Figure 7: Correlation Between ELA GPC B Parameters

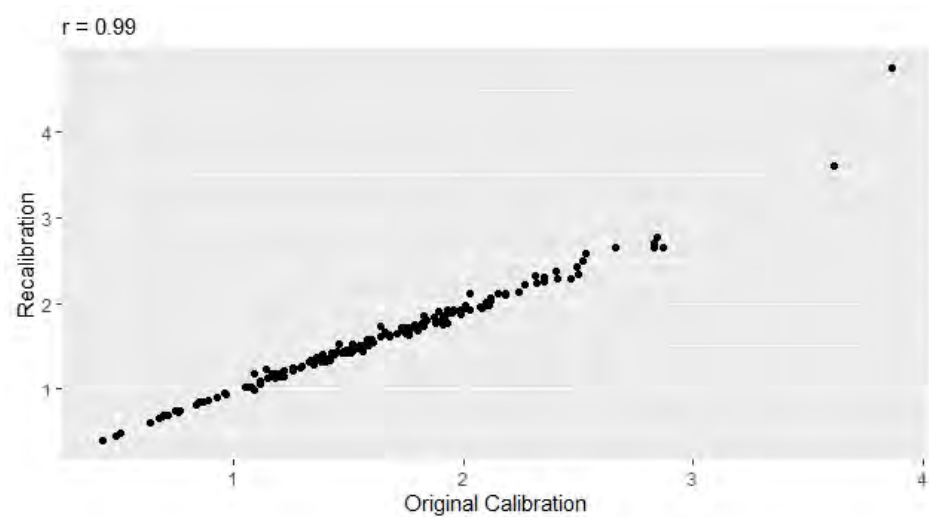


Figure 8: Correlation Between Math 2PL A Parameters

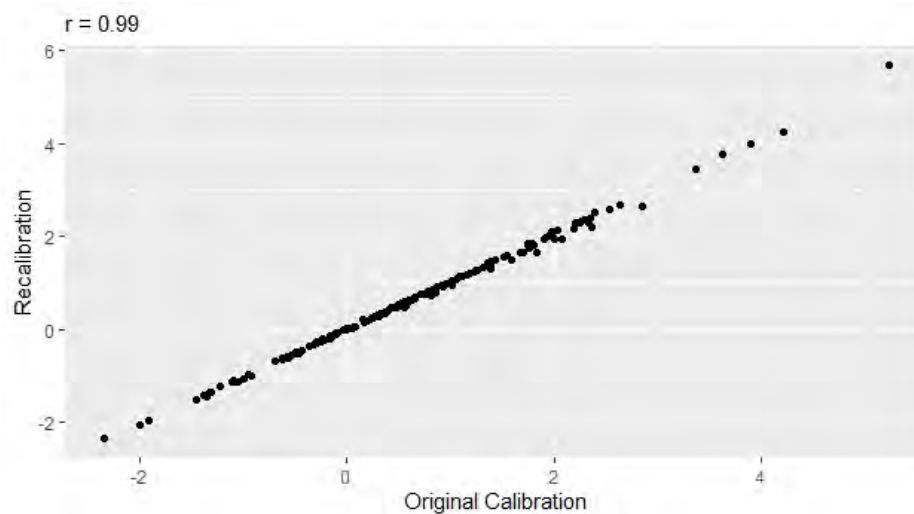


Figure 9: Correlation Between Math 2PL B Parameters

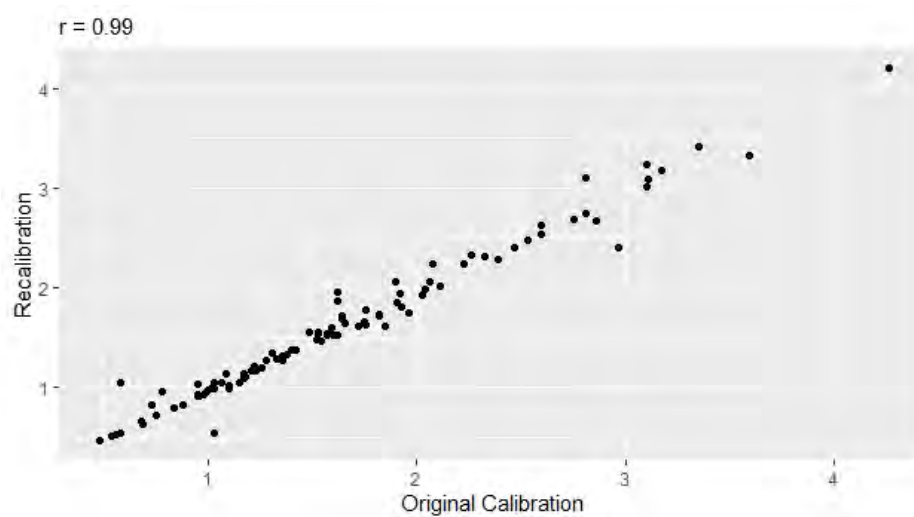


Figure 10: Correlation Between Math 3PL A Parameters

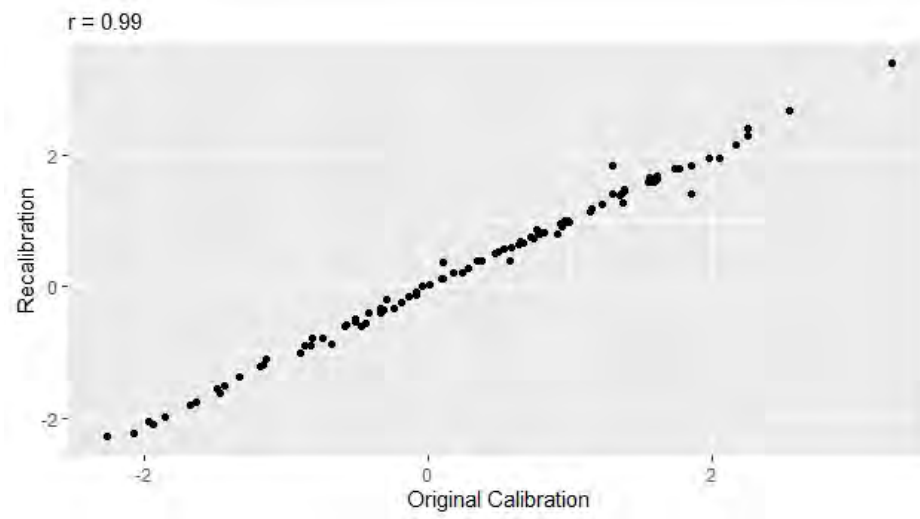


Figure 11: Correlation Between Math 3PL B Parameters

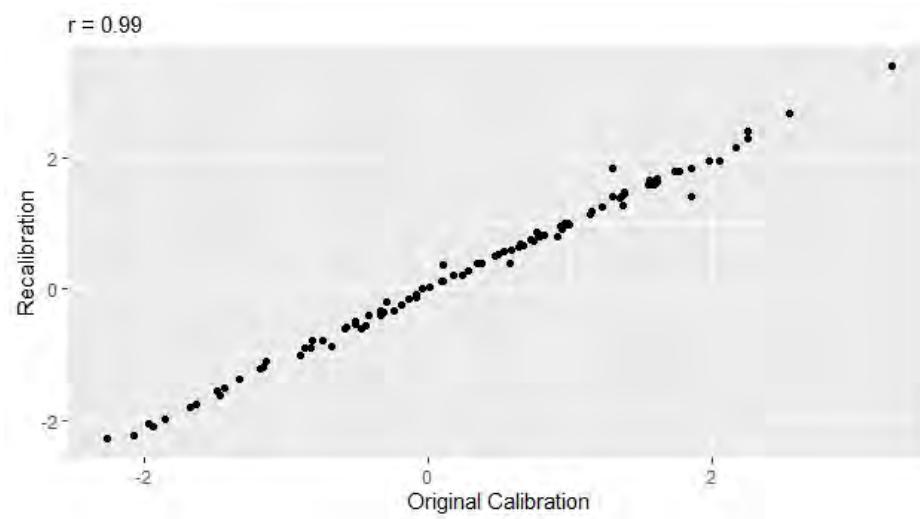


Figure 12: Correlation Between Math 3PL C Parameters

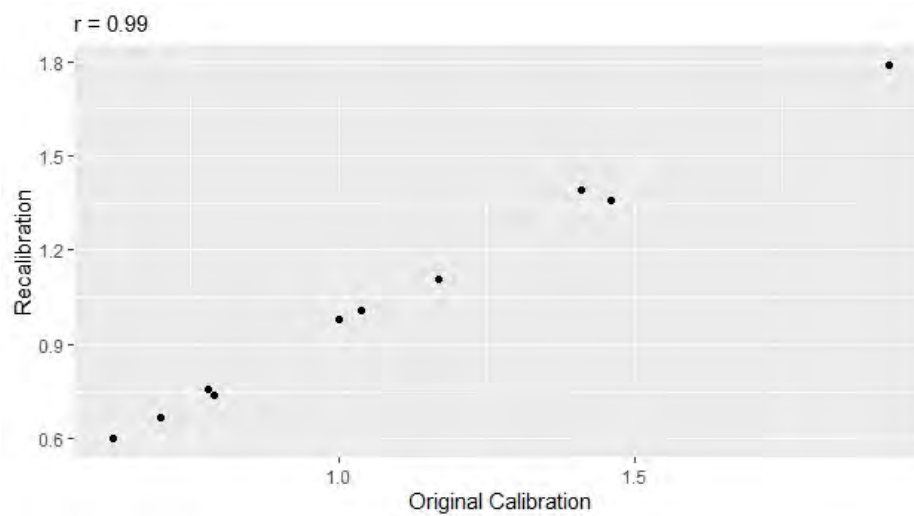


Figure 13: Correlation Between Math GPC A Parameters

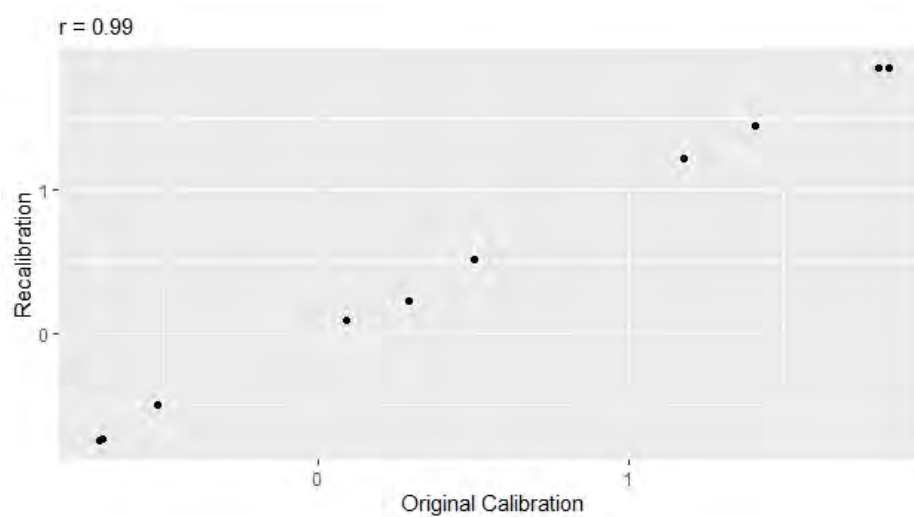


Figure 14: Correlation Between Math GPC B Parameters

Conclusions

A concern that has been raised regarding the purging of all operational items in flagged content standards from item calibration is the disruption in content representativeness across the ICCR item bank. In their discussion of assessments in the era of COVID-19, Keng, Boyer, and Marion (2020) recommended that states continue to maintain their current test design as the removal of content standards can lead to instability in the measurement scale and a disturbance to the achievement trend lines. With the high number of content standards flagged with misfitting items in the current study, their removal would create a significant change in this regard.

It is important to note that there were some limitations in the present investigation. Not all items were eligible for field testing across all states, and several items were removed from the study due to small sample sizes. It was also possible that an item might fit better in one state than another. This was not examined in the current study as global fit indices were the primary concern.

However, we found that the inclusion of the flagged anchor and field test items had a minimal impact on item parameter estimation, with item parameters from both conditions shown to be highly correlated and demonstrating comparable means across scoring models. Additionally, the differences between item parameter SEEs between the two calibrations were found to be low, with none of them meeting the significance criteria of an absolute value of 1. It was also found that the removal of these items in the second calibration affected calibration convergence for ELA items across several grades. With this, the decision was made to retain the field test items from the spring 2021 administration and from the original calibration.

References

- Boyer, M., Dadey, N., Keng, L. (2020). Statewide summative assessment in spring 2021: A workbook to support planning and decision-making. Dover, NH: National Center for the Improvement of Educational Assessment.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.
- Guo, H., Robin, F., & Dorans, N. (2017). Detecting item drift in large-scale testing. *Journal of Educational Measurement*, 54, 265 – 284.
- Guo, H., Zheng, Y., & Chang, H. (2015). A stepwise test characteristic curve method to detect item parameter drift. *Journal of Educational Measurement*, 52, 280 – 300.
- Keng, L, Boyer, M., Marion, S. (2020). Into the unknown: Assessments in Spring 2021. *Educational Measurement: Issues and Practice*, 39, 53 – 59.

REPORT

**Overall Results of an Alignment Analysis of the
Cambium Statewide Science Assessment Item Bank
for States Sharing Science Content
with Corresponding Grade Level Next Generation Science
Standards (NGSS) Performance Expectations (PEs)**

**Sara C. Christopherson and Norman L. Webb
May 26, 2020**

Wisconsin Center for Educational Products and Services
Matt Messinger, Executive Director
University of Wisconsin Research Park
510 Charmany Drive, Suite 269
Madison, WI 53719

Acknowledgements

Elementary School Science:

Anna Hardway	Group Leader (3-5) Texas
<i>Panelists</i>	
Theresa White	Vermont
Taya Trent	West Virginia
Diane Hart	Connecticut
Joe Evans	West Virginia
Liz David	Wyoming
Tamre Mockus	Connecticut
Cathy Jimenez Long	Rhode Island
Ericka Swett	New Hampshire
Sharon Angal	Oregon
Zachary Sallade	West Virginia

Middle School Science:

Karen Mesmer	Group Leader (6-8) Wisconsin
John Putnam	Group Leader (6-8) West Virginia
<i>Panelists</i>	
Melyssa Ferro	Idaho
Alvin Lin	Hawai'i
Julie Caulkins	Wyoming
Ricky Scott	Utah
Phyllis Samuel	West Virginia
Tabitha Secretario	Hawai'i
Ben DeCarlow	Oregon
Donna Casanova	Rhode Island
Jennifer Blay	Vermont

High School Science:

Tim Cox	Group Leader (9-12) Wisconsin
Aaron Allen	Group Leader (9-12) West Virginia
<i>Panelists</i>	
Josh Jordan	Oregon
Diane Blake	Connecticut
Nathan Reutter	Vermont
Tiffany Risch	Rhode Island
Joanne Goelzer	New Hampshire
Emilio Macalalad	Hawai'i

This independent third-party analysis was conducted for the states of Connecticut, Hawai'i, Idaho, New Hampshire, Oregon, Rhode Island, Utah, Vermont, West Virginia, and Wyoming and coordinated and funded by Cambium Assessments (formerly American Institutes for Research – AIR). Toni Deoudes was the main contact for communication with Cambium Assessment. A Working Group comprised of state leadership also contributed to study planning and coordination.

Table of Contents

Table of Contents

Executive Summary	1
Introduction and Methodology.....	6
Overview of Item Bank, Targeted Standards, and Scoring Assertions	9
Cognitive Engagement (Cognitive Complexity) Evaluative Tool	14
Trial Run Study	15
Panelists	19
Training and Coding.....	20
Data Analysis	26
Alignment Criteria Used for This Analysis.....	28
Reporting Categories and Standards.....	32
Use of Phenomena	32
Dimensionality/Structure of Knowledge Comparability	32
Categorical Concurrence	33
Consistency of Cognitive Engagement (Category of Engagement).....	33
Range of Knowledge Correspondence	35
Balance of Representation.....	36
Relationship of Scoring Assertions with Student Interactions	36
Relationship of Scoring Assertions with PEs	36
Source of Challenge	37
Summary Findings: Shared Science Assessment Item Bank.....	38
Standards.....	39
Consistency of Cognitive Engagement	41
Use of Phenomena	43
Categorical Concurrence	43
Range of Knowledge Correspondence	44
Balance of Representation.....	45
Relationship of Scoring Assertions with Student Interactions.....	45
Relationship of Scoring Assertions with PEs	46
Relationship of Scoring Assertions with PEs	46
Items Flagged for Revisions or Removal	46
Source of Challenge issues and Reviewer Comments	48
Reliability Among Panelists.....	49
Conclusions	52
References.....	59

Appendix A: Group Consensus Category of Engagement Values for the Next Generation Science Standards' Performance Expectations (PEs)

Appendix B: Data Analysis Tables for Each Sample Test Form and Item Batch

Appendix C: Items Flagged by Panelists for Revision [Redacted for Public Release]

Appendix D: Panelists' Notes [Redacted for Public Release]

Appendix E: Debriefing Summary Notes [Redacted for Public Release]

Appendix F: Category of Engagement Definitions for Content Area

Appendix G: Materials Provided to Panelists

Executive Summary

The WebbAlign team of the non-profit Wisconsin Center for Education Products and Services (WCEPS) facilitated a study in July 2019 to evaluate the alignment of a Shared Science Assessment Item Bank with the Next Generation Science Standards' (NGSS) Performance Expectations (PEs). The item bank is managed and owned in part by Cambium Assessment (CAI), (formerly known as American Institutes for Research; AIR) and is shared by multiple states, 10 of which were involved in the 2019 study. Each state uses its own assessment blueprint and a particular state-vetted subset of items from the shared item bank. Some states' science standards include slight adjustments from the wording or scope of the NGSS. This report describes the overall results of the alignment analysis of the item bank, applicable to all states. Separate state-specific reports provide additional detail relevant to each state.

The alignment analysis was designed to yield the evidence that could (as appropriate, pending results) substantiate state claims about what the assessments measure as well as state interpretations of student scores in relation to the NGSS. This includes the evidence required for submission to federal peer review and reflects the input, discussions, and decisions of participating states, input from the VT/RI (MSSA) Technical Advisory Committee, and the takeaways from a small-scale trial of alignment methodologies using item clusters from the item bank. The methodologies used for the alignment analyses synthesize current thinking and specific considerations as relates to NGSS alignment as well as core tenets of evaluations of alignment of standards and assessments that meet U.S. Department of Education expectations.

The in-person content alignment analyses were conducted in Denver, CO on July 15-19, 2019 with panelists from all 10 states. Ten to twelve educators for each grade band (elementary, middle school, and high school) were split into two panels of five and/or six. Panels were expected to have appropriate state representation as well as education and science content/discipline experience and expertise. The first day of the meeting was dedicated to large-group and small-group training and practice, as well as a thorough content analysis of the standards by grade band panels to promote a shared interpretation of their meaning. On days two through four, panelists completed a content analysis of all operational assessment stand-alone items and item clusters (available at the time of the study) within the Shared Science Assessment Item Bank. Group leaders had content area and NGSS expertise as well as previous experience with WebbAlign alignment analyses. Group leaders worked in advance with the Study Director to prepare for study facilitation.

Research questions, alignment criteria, and acceptable cutoff levels for these criteria as relate to corresponding science standards and assessments were determined through discussion with state officials, and grounded in analyses of test purpose, construct, and blueprint. For NGSS assessments, judgements of acceptability of alignment must also relate to and be informed by the specifics of the structure of the standards and the assessment design and construct. For example, states expected item clusters to require students to engage with all three dimensions of the standards while stand-alone items could require engagement with two or three dimensions of the standards. State officials were asked to confirm the full set of criteria used and review the acceptable levels proposed for each alignment criterion. States were provided the opportunity to make modifications if warranted. The data collection and reporting for the analyses used the finalized alignment criteria and corresponding cutoffs, appropriate for the context of the states' intents and for the particular context of alignment of science assessments with corresponding NGSS PEs.

The study addressed four key research questions:

1. To what extent do the stand-alone items and item clusters satisfy the measurement target claims (PE and scoring assertions) identified in the CAI metadata?
 - a. To what extent does an independent expert panel agree that a student's correct response allows for a reasonable inference about the student's proficiency as relates to the three-dimensional expectations within the identified PE?
 - b. To what extent does an independent expert panel agree that the explicit inferences about student performance stated in the scoring assertions can reasonably be made based on student responses to a stand-alone item or item cluster?
 - c. To what extent does an independent expert panel agree that the explicit inferences about student performance stated in the scoring assertions reflect the states' measurement target claims (PE) identified in the CAI metadata?
2. What Category of Engagement (cognitive complexity) is required for successful completion of each interaction within a stand-alone item or item cluster and how does this compare with the Category of Engagement assigned to the corresponding PE?
3. To what extent do the stand-alone items and item clusters satisfy the claim that the assessment is phenomenon-based?

4. To what extent are state-specific assessment programs likely to generate test events that are aligned with corresponding grade-level academic standards, considering depth and breadth (specified in ESSA) as well as other alignment criteria?
 - a. Do the test blueprints and other relevant test specifications and documentation reflect appropriate design to support potential alignment of test events with corresponding grade-level academic standards?
 - b. Do the available aggregate data for recently administered test events in each state provide evidence that the algorithm and blueprints yielded test forms as expected?
 - c. To what degree are actual test events for each state (if available) aligned with corresponding grade-level academic standards for each state?

The overall study was crafted to allow for the potential to build a logic argument for the capacity for alignment of all test events generated by the Shared Science Assessment Item Bank with corresponding state standards, as appropriate, based on results. As such, the study was designed to generate multiple lines of evidence that could be used to support a claim that the item bank has the capacity to yield aligned test forms for each state. These lines of evidence, along with the resulting claim, stated in the positive, would be:

- If an independent content alignment analysis of all stand-alone items and item clusters from the item bank shows that the items/clusters are appropriate as relates to intended claims/inferences,
- and if a state's test blueprints and item selection algorithm are generating test events/forms as intended (based on data from all administered test events within that state),
- then it is possible to make an argument for the capacity of alignment for all test events resulting from the state's summative science assessment program that use items from the shared item bank.

This argument could be further supported if an independent content analysis of actual sample test events for a state shows acceptable alignment to corresponding standards per agreed-upon criteria. At the time of the in-person content analysis, sample test events were available for Connecticut, New Hampshire, and Oregon. Other states did not have sample test events available. In many cases, sample test events from other states with very similar blueprints and item banks may serve as evidence by proxy.

Nine alignment criteria, listed below, received major attention. These criteria were identified and confirmed through discussions with the state representatives who formed a collective decision-making body known as the Working Group.

1. **Use of Phenomena:** Each stand-alone item and item cluster is expected to be grounded in a stimulus that meets the test development criteria for a phenomenon. Items/clusters are expected to require students to engage multiple dimensions of the PEs (“use science”) to make sense of those phenomena.
2. **Categorical Concurrence:** Test events are expected to yield sufficient evidence to make inferences about student knowledge, skills, and abilities (KSAs) as relates to each reporting category.
3. **Dimensionality (Structure of Knowledge):** All item clusters are expected to require students to demonstrate integrated engagement with the three dimensions of Science and Engineering Practices (SEPs), Disciplinary Core Ideas (DCIs), and Crosscutting Concepts (CCCs) specified in the targeted PE. All stand-alone items are expected to require students to demonstrate integrated engagement with two or three of the dimensions specified in the targeted PE.
4. **Consistency of Cognitive Engagement:** The assessment is expected to elicit work that is as cognitively demanding as the expectations in the PEs.
5. **Range of Knowledge Correspondence (Individual):** Test events are expected to assess an appropriate breadth of the standards. For individual students, assessed PEs are sampled across topics within each reporting category.
6. **Range of Knowledge Correspondence (Population):** At least 90% of PEs within a grade band have the potential to be assessed across the student population. State-specific claims are consistent with aggregate data from all administered test events in the state in conjunction with results from an independent analysis of vendor metadata.
7. **Balance of Representation:** No PE is targeted more than once on any single test event.
8. **Relationship of Scoring Assertions with Student Interactions:** In aggregate, the scoring assertions for an item/item cluster appropriately represent the inferences about student knowledge, skills, and abilities that can be made based on successful interactions with an item/cluster.
9. **Relationship of Scoring Assertions with PEs:** In aggregate, the scoring assertions for an item/item cluster appropriately represent the three-dimensional expectations of the targeted PE.

Study results suggest that the overall Shared Science Assessment Item Bank for both the elementary and middle school has the capacity to fully meet all alignment criteria used in this study and itemized above. The Shared Science Assessment Item Bank for high school has the capacity to weakly meet the criterion of Range of Knowledge Correspondence (Population) and the capacity to fully meet all other alignment criteria used in this study and itemized above. The criteria of Categorical Concurrence, Range of Knowledge Correspondence (Individual), and Balance of Representation are addressed primarily in the state-specific reports, considering state-specific documentation, test events, and data. In general, full or acceptable alignment was found between the NGSS PEs and the assessment items and item clusters for all grade bands although specific items were identified that did not meet one or more expectations, and warrant revisions or removal (details provided within the Findings section).

The only alignment weakness identified was that the high school item pool did not meet states' expectations to have the capacity to address at least 90% of the corresponding PEs. This issue could be fully resolved with the addition of at least six items to the high school item bank. At least five of these items would need to address unrepresented PEs within the Physical Science domain.

Even for items that panelists agreed met alignment expectations, many editorial suggestions were made to correct errors found in text and graphics, improve clarity, and/or address scientific inaccuracy. This extent of editorial issues is typically not observed in an operational assessment and included many issues that could affect student scores. Overall, however, panelists found that items and item clusters were meeting state expectations that assessment tasks required integrated engagement with at least two (stand-alone items) or three (item clusters) dimensions of SEPs, DCIs, and CCCs specified in the targeted PE in order to make sense of a phenomenon. With just a very few exceptions, items required student cognitive engagement consistent with the expectations of the standards. With the sole exception of the high school item pool for Physical Science, there was one or more item(s) or item cluster(s) that represented all or all but one or two PEs within each grade band and domain. Item were spread across the domains of Physical, Life, and Earth and Space Science, with no PE(s) overemphasized in the item bank. Overall, panelists found that the large majority of scoring assertions reasonably reflected inferences that could be made based on student interactions and corresponded to the expectations within the targeted PE. One panelist summarized "Overall, the items seem strong and do a commendable job of assessing proficiency as it relates to the 3D standards."

Introduction and Methodology

The alignment of expectations for student learning with assessments for measuring students' attainment of these expectations is an essential attribute for an effective standards-based education system. The critical role of alignment in the success of Framework-influenced science standards, including but not limited to the Next Generation Science Standards (NGSS) was recognized and called out by the committee in the very first chapter of A Framework for K-12 Science Education:

"The committee recognizes that the framework and subsequent standards will not lead to improvements in K-12 science education unless the other components of the system – curriculum, instruction, PD, and assessment – also change so they are aligned with the framework's vision." (NRC, 2012)

In the context of statewide summative assessments, alignment is defined as the degree to which expectations (standards) and assessments are in agreement and serve in conjunction with one another to guide an education system toward students learning what they are expected to know and do. As such, alignment is a quality of the relationship between expectations and assessments and not an attribute solely of either of these two system components. Alignment describes the match between expectations and an assessment that can be legitimately improved by changing either student expectations or the assessments. As a relationship between two or more system components, alignment is typically determined by using, at minimum, multiple criteria described in detail in a National Institute for Science Education (NISE) research monograph, *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education* (Webb, 1997). The corresponding Webb methodology used to evaluate alignment has been refined and improved over the last 20 years, yielding a flexible, adaptable, effective, and efficient analytical approach. Some version of this alignment methodology has been used to analyze curriculum standards and assessments in nearly all states to satisfy or to prepare to satisfy the Title I compliance as required by the United States Department of Education (USDE). Modified and/or expanded versions of this alignment methodology have been used for studies of multi-dimensional assessments, computer adaptive tests (CATs), interim assessments, alternate assessments, for studies intended to inform vendor internal continuous improvement, and other purposes. Evidence of alignment is one of multiple lines of evidence that are necessary for building a validity argument that student scores from an assessment can reasonably yield the intended inferences.

The study detailed in this report was coordinated and facilitated by the WebbAlign program, which works directly with Dr. Norman Webb, and operates out of the Wisconsin Center for Education Products and Services (WCEPS), a non-profit organization that strives to extend the reach of innovations developed at the University of Wisconsin-Madison's Wisconsin Center for Education Research (WCER). Sara Christopherson led the study in collaboration with Norman Webb. Sara has participated in and led Webb alignment studies since 2005 for state departments of education in nearly 30 states as well as for assessment vendors and other entities. Sara's practical and academic experience with K-12 and tertiary science education efforts over the last two decades, leading up to and including NGSS, also informed study design.

An item- and item-cluster-level content alignment analysis was completed in July 2019 to evaluate the alignment of the Shared Science Assessment Item Bank with the Next Generation Science Standards' (NGSS) Performance Expectations (PEs). The item bank is managed and owned in part by Cambium Assessment (CAI) (formerly American Institutes for Research; AIR) and is shared by multiple states, 10 of which were involved in the 2019 study. Participating states agreed to share items through a Memorandum of Understanding (MOU) that details a commitment to share content, leadership, ideas, and methods. Each state uses its own assessment blueprint and a particular state-vetted subset of items from the shared item bank. Some states' science standards include adjustments from the wording or scope of the NGSS PEs. This report describes the overall results of the alignment analysis of the item bank, applicable to all states. Separate state-specific reports provide additional detail relevant to each state.

The July 2019 study was structured to answer four key research questions:

1. To what extent do the stand-alone items and item clusters satisfy the measurement target claims (PE and scoring assertions) identified in the CAI metadata?
 - a. To what extent does an independent expert panel agree that a student's correct response allows for a reasonable inference about the student's proficiency as relates to the three-dimensional expectations within the identified PE?
 - b. To what extent does an independent expert panel agree that the explicit inferences about student performance stated in the scoring assertions can reasonably be made based on student responses to a stand-alone item or item cluster?
 - c. To what extent does an independent expert panel agree that the explicit inferences about student performance stated in the scoring assertions reflect the states' measurement target claims identified in the CAI metadata?

2. What Category of Engagement (cognitive complexity) is required for successful completion of each interaction within a stand-alone item or item cluster and how does this compare with the Category of Engagement assigned to the corresponding PE?
3. To what extent do the stand-alone items and item clusters satisfy the claim that the assessment is phenomenon-based?
4. To what extent are state-specific assessment programs likely to generate test events that are aligned with corresponding grade-level academic standards, considering depth and breadth (specified in ESSA) as well as other alignment criteria?
 - a. Do the test blueprints and other relevant test specifications and documentation reflect appropriate design to support potential alignment of test events with corresponding grade-level academic standards?
 - b. Do the available aggregate data for recently administered test events in each state provide evidence that the algorithm and blueprints yielded test forms as expected?
 - c. To what degree are actual test events for each state (if available) aligned with corresponding grade-level academic standards for each state?

Research questions 1-3 are addressed in this report. Research question 4 is addressed in the state-specific reports.

The overall study was designed to generate multiple lines of evidence that would allow for the potential to build a logic argument for the capacity for alignment of all test events generated by the Shared Science Assessment Item Bank with corresponding state standards, as appropriate, based on results. In other words, the findings from the item-level analyses, along with information about test blueprints and data from all administered test events, can be used to make inferences about the capacity of each state's assessment system, overall, to yield aligned test events. The resulting logic argument, stated in the positive, would be:

- If an independent content alignment analysis of all stand-alone items and item clusters from the item bank shows that the items/clusters are appropriate as relates to intended claims/inferences,
- and if a state's test blueprints are appropriately designed and, along with the item selection algorithm, are generating test events as intended (based on data from all administered test events within that state),
- then it is possible to make an argument for the capacity of alignment for all test events resulting from the state's summative science assessment program that use items from the Shared Science Assessment Item Bank.

This argument could be further supported if an independent content analysis of actual sample test events for a state shows acceptable alignment to corresponding standards per agreed-upon criteria. At the time of the in-person content analysis, sample test events were available for Connecticut, New Hampshire, and Oregon. Other states did not have sample test events available. In many cases, sample test events from other states with very similar blueprints and item banks may serve as evidence by proxy. This report details the summary findings for all grades and domains and includes suggestions for areas in which improvements are needed. Throughout this report, “item” may refer to both stand-alone items and to item clusters. While both may require multiple student interactions, the unit of analysis is the overall item or item cluster.

Overview of Item Bank, Targeted Standards, and Scoring Assertions

The overall item bank analyzed in this study consisted of AIRCore items owned by CAI as well as items owned by particular states. Regardless of ownership, all items followed the same test development and review processes. The item bank includes stand-alone items and item clusters (see **Figure 1**) for elementary, middle, and high school sciences that are grounded in NGSS performance expectations (PEs). Results as relates to any PEs unique to a state are addressed in that state’s state-specific report.

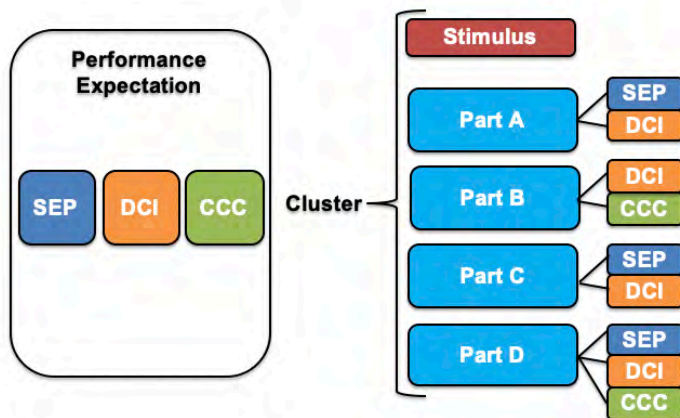
Each stand-alone item and each item cluster within the item bank was designed to address a single performance expectation. Performance expectations are intended to provide “clear and specific targets...for assessment” and describe “what students should be able to do **after** instruction” (NGSS, 2013). Each PE is written as a statement that interweaves the three dimensions of the standards: Disciplinary Core Ideas (DCIs), Science and Engineering Practices (SEPs), and Crosscutting Concepts (CCCs). Some participating states adapted the PEs such that the wording is slightly adjusted or additional expectations are included. Any adjustments to the NGSS PEs that affect the DCI, SEP, and/or CCC is addressed in the state-specific reports. Slight adjustments to wording that did not affect the DCI, SEP, or CCC were considered insignificant. For example, the NGSS PE 5-PS2-1 refers to gravitational force being “directed down” and in the clarification statement defines “down” as “the direction that points toward the center of the spherical Earth.” Rhode Island and Vermont edited this PE to use the phrase “directed toward the center of the Earth” rather than using “down.” Because the three dimensions of the PE and overall meaning of the expectation remain the same, this type of editorial change was considered a clarification or preference.

Both stand-alone items and item clusters are intended to be based on a specific real-world scenario and focused enough to require students’ application of a Science and Engineering Practice (SEP) in the context of a Disciplinary Core Idea (DCI) and Crosscutting Concept (CCC) as intended by the PE in order to

make sense of the phenomenon presented. (States considered it acceptable for stand-alone items to require application of two of the three dimensions.) Item clusters are multi-part items that include an extensive scenario, typically involving text, illustrations, data shown in a variety of formats, short animations and other features. Item clusters have between several and up to around 20 different student interactions. Item clusters are typically, but not always, presented on-screen via two panels. One panel provides the stimulus. The other panel contains the instructions, prompts, and answer spaces for the student interactions.

Figure 1 shows the structural composition of an item cluster. As shown, some parts of an item cluster include student interactions that are intended to draw on two dimensions of the standards, while others are intended to draw on all three. Each part of an item cluster may include multiple student interactions. Stand-alone items present a more concise scenario and include, at most, several student interactions. Assessments are administered online.

Figure 1. Structure of Item Clusters in the Shared Science Assessment Item Bank



Source: AIR, 2018

Each stand-alone item and item cluster is scored with a set of binary (true/false) narrative scoring assertions, which constitute the scoring rationales for items. Per CAI, each assertion describes a piece of content knowledge, skill, or ability (KSA) that is related to the targeted PE and that the student is expected to have demonstrated by successful interaction with the item. In general, an assertion states the student's action(s) within the item that provide(s) evidence for the corresponding inference about student KSAs. Per assessment design, the number of scoring assertions for an item varies, depending on the evidence that an item can yield based on a student's response. Stand-alone items have an average of two scoring assertions and up to seven scoring assertions, maximum.

Item clusters have an average of 10 scoring assertions and up to 20 scoring assertions, maximum.

Items included many different types of interactions. The variety of types of student interactions that could be used in each part of a stand-alone item and an item cluster is shown in **Table 1**. All interactions were machine scorable.

Table 1. Types of Student Interactions Used in Stand-Alone Items and Item Clusters within the Shared Science Assessment Item Bank

Interaction Type	Associated Sub-Types	Description
Choice	Multiple-Choice	Traditional multiple-choice interaction allows the student to select a single option from a list of possible answer options.
	Multi-Select	Traditional multi-select interaction (checkboxes) allows students to select one or more options from a list of possible answer choices.
Text Entry	Simple Text Entry	Students type a response in a text box.
	Embedded Text Entry	Students type their response in one or more text boxes that are embedded in a section of read-only text.
Table	Table Match	Interaction allows students to check a box to indicate if the information from a column header matches information from a row header.
	Table Input	Interaction solicits a student to complete tabular data.
Edit Task	Edit Task	A student clicks a word and replaces it with another word that they type to revise a sentence.
	Edit Task with Choice	A student clicks a word or phrase and chooses the replacement from a number of options.
	Edit Task Inline Choice	Drop-down menus are placed through the text, and a student chooses the right option to complete the text.
Hot Text	Selectable	Selectable hot text interactions require students to select one or more text elements in the response area.
	Re-orderable	Re-orderable hot text interactions require students to click and drag hot text elements into a different order.
	Drag-from-Palette	Drag-from-Palette hot text interactions require students to drag elements from a palette into the available blank table cells or "gaps" (text boxes) in the response area.
	Custom	Custom hot text interactions combine the functionality of the other hot text interaction sub-types. Students responding to a custom hot text interaction may need to select text elements, rearrange text elements, and/or drag text elements from a palette to blank table cells or drop targets in the response area.

Table 1 Cont'd. Types of Student Interactions Used in Stand-Alone Items and Item Clusters within the Shared Science Assessment Item Bank

Interaction Type	Associated Sub-Types	Description
Equation	n/a	Equation interactions require students to enter a response into input boxes. These boxes may stand alone, or they may be in line with text or embedded in a table. The equation interaction may have an on-screen keypad which may consist of special mathematics characters. Students may also enter their response via a physical keyboard.
Grid	Grid	Grid interactions require students to enter a response by interacting with a grid area in the answer space. The student may be required to draw a line or shape, plot a point, or create a graph. The student may also drag and drop or click on selectable hot spots.
	Hot Spot	Hot spot interaction sub-types allow you to create grid interactions with specific hot spot functionality. These interactions require students to select hot spot regions in the grid area.
	Graphic Gap Match	Graphic gap match interactions allow you to create grid interactions with specific drag-and-drop functionality. These interactions require students to drag image objects from a palette to specified regions (gaps) in the grid area.
Simulation	n/a	Simulation interactions allow the student to investigate a phenomenon by selecting variables to get output data. Some simulations are accompanied by animations.

Source: CT Technical Report, Vol. 2 (2020)

The total operational item count as of Spring, 2019 and as reported in the CAI Technical Report is shown by grade band in **Table 2**. Each state works with CAI leadership to develop statewide summative assessment programs that use items drawn from the Shared Science Assessment Item Bank. Test events are administered online and use linear-on-the-fly (LOFT) item delivery in which the item selection algorithm chooses items based solely on content value toward blueprint fulfillment. In other words, each item is selected based on its contribution to meeting the blueprint specifications, given the items that have already been administered.

For most states, blueprints separate items by the domains (reporting categories) of Life Sciences, Physical Sciences, and Earth and Space Sciences. Each domain is further divided into sub-domains according to the NGSS DCI arrangement of the PEs. Blueprints specify the length of the test and the minimum and maximum number of items that can be included on a test event by DCI organization of the PEs per domain. For all states, blueprints specify that each PE is represented on a test event by no more than one item cluster or stand-alone item. In general, blueprints specify that a test event can include no

more than one item cluster or two stand-alone items that target PEs within the same DCI sub-domain.

Table 2. Shared Science Assessment Item Bank Operational Items, Spring 2019

Grade Band	Item Type	Total Operational Items as of Spring 2019
Elementary School	Cluster	76
	Stand-Alone	59
Middle School	Cluster	137
	Stand-Alone	57
High School	Cluster	66
	Stand-Alone	56
Total		451

Source: CT Technical Report, Vol. 2 (2020)

Panelists analyzed the full set of operational items and item clusters available at the time of the study to allow for the most direct inferences and conclusions to be made about alignment of test events with standards when extrapolating findings from the item bank review to specific states. Some items were removed from the item bank between the Spring 2019 administration and the alignment study. The total number of items reviewed in the July 2019 study and included in the results within this common report is 440. For each grade band, CAI provided a randomly selected sample test event from New Hampshire. The New Hampshire test events should be substantially representative of Connecticut, Hawaii (for elementary and middle school only), Oregon, Rhode Island, Vermont, and West Virginia. All of these states are NGSS adopters, meaning that they use the exact language of the NGSS PEs. They also use very similar blueprints that draw from a very similar item bank. For middle school and high school, a randomly selected test event from each of Connecticut and Oregon was also provided for analysis. The remaining items for each grade band were organized into “batches” for review. At the time of the study, some states did not yet have operational blueprints or finalized item banks. As such, sample test events were not available for all states. State-specific considerations are addressed in the state-specific reports.

NGSS documentation specifies that “[s]ome states consider th[e] performance expectations alone to be ‘the standards,’ while other states also include the content of the three foundation boxes and connections to be included in ‘the standard’” (NGSS, 2013). The foundation boxes are shown beneath a PE and provide additional information about the specific elements of the three dimensions within a particular PE. Appendices E, F, and G of the NGSS provide yet more detail on each of the three dimensions. The study director sought clarification through discussions with stakeholders on states’ interpretation of the

PE as “the standard,” e.g. as relates to the consideration of the information within the foundation boxes for each dimension and as relates to consideration of the corresponding NGSS appendices for each dimension. Perspectives were reconciled as needed to ensure clarity on the specific information that states expected to be used by educators (and correspondingly, in the alignment analysis) for interpretation of each PE. The study director also sought clarification on stakeholder interpretation of the PE as the statewide summative assessment target, including which attributes of the PE are expected to be assessed (or not) in this particular assessment context. All states expected use of the foundation boxes, at least in the context of the assessment development and analysis. Panelists evaluated the specific elements within the foundation boxes to evaluate the degree of alignment (except when considering Utah standards).

Cognitive Engagement (Cognitive Complexity) Evaluative Tool

A Framework for K-12 Science Education and the resulting NGSS both emphasize a conceptual shift in science standards, related to the complexity of student engagement with science concepts and scientific thinking (NGSS Appendix A, Conceptual Shift #4). The alignment analysis examined if and in what ways different types of student cognitive engagement (cognitive complexity) were being interpreted both in the expectations and the assessment. To achieve this, the Categories of Engagement evaluative tool for differentiating categories of complexity/types of engagement was used in this study. The development of this tool was informed by input from across the science education field, solicited during a 2-day in-person summit held in May 2018 as well as via follow-up and feedback on initial drafts. Stakeholders and thought leaders at the summit included NGSS and Framework writers and committee members, as well as practitioners and measurement experts. The tool is intended to be used equivalently to DOK, as a language system and conceptual model, to evaluate and differentiate between and among different types of cognitive engagement, but with important revisions to reflect current science standards such as NGSS, as well as other Framework-based and three-dimensional standards. The draft tool focuses on the type of student interaction with a phenomenon, parses out components of sense-making and knowledge-in-use, relates directly to concepts of transfer, reasoning with evidence, and other NGSS-specific expectations, and explicitly connects back to three-dimensional and Framework-based expectations. The tool differentiates complexity of engagement from task difficulty. It also differentiates the concepts of complexity of engagement from dimensionality as well as sophistication of thinking or performance.

The drafted definitions of the Categories of Engagement were field tested in a trial-run study and with classroom educators in advance of the alignment institute. Feedback was then incorporated into the definitions. In preparation for the alignment analysis, the tool was presented to the state Working Group for consideration, discussed, and approved for use. Definitions for each Category

(complexity level) are included within **Appendix F** along with additional information about the development of the tool.

Trial Run Study

The evaluative steps used in the July 2019 study were informed by a trial run of the process conducted over June 4-5, 2019. To recruit panelists, an email invitation was shared with Kevin Anderson (Wisconsin Department of Public Instruction; President of the Council of State Science Supervisors), who broadcast the invite via the Wisconsin Science Education Leadership Association (WSELA) listserv. Only regional educators were invited due to time constraints. Kevin Anderson, along with 11 regional classroom educators with NGSS experience, agreed to participate. The classroom educators all submitted resumes that met expectations for experience and expertise. Ten educators participated in the first day of the meetings. One educator was not able to participate in the second day of the meetings due to unexpected conflicts. Overall, nine expert educators participated as panelists in the full trial. All states were invited to listen in on the trial run via online conferencing.

Participating panelists provided feedback based on their experience in the trial run. State Working Group members also provided feedback based on their observations of the trial run. Consensus on several themes emerged from the trial run that motivated adjustments in the plans for several evaluative steps to be used in the July 2019 alignment analysis. Overall, modifications to the methodology were made to clarify and streamline instructions and processes, including reconsidering the grain-size of some components of the analysis. The purpose of these adjustments was to limit cognitive overload on panelists and to allow for the potential to realistically complete the analysis in a reasonable amount of time, yielding the necessary data but tailored to eliminate any excess. For example, in the trial run, panelists were asked to individually code each student interaction within an item cluster and to individually code each scoring assertion for an item. This resulted in valuable feedback about individual scoring assertions but took an excessive amount of time and was determined to be too fine-grain of an analysis for the purposes of the alignment study. Although all subcomponents must be evaluated, it was decided that panelists would record data only at the larger item cluster level and holistic (aggregate) scoring assertion level. This grain-size is consistent with the units of analysis identified in the assessment specifications: an item/item cluster is intended to target a PE, and the scoring assertions, in aggregate, are intended to reflect the inferences about student KSAs as relates to the PE. This grain size is also consistent with the purpose of the content alignment analysis, which is to judge the overall extent to which the scores from the assessments can reasonably be used to make inferences about student learning as relates to the PEs. A content alignment analysis is a fairly high-level analysis of the extent to which an assessment is measuring what it is intended to measure.

Another key shift in the planned coding processes related to the method for determining which PE, if any, should be coded to an item or item cluster. In the trial run, panelists individually coded items and item clusters, completing the full set of evaluative steps before any adjudication. Completing coding for a single item cluster could take 20-30 minutes, sometimes even more. Particularly for item clusters, the panelists found that the student interactions often engaged more than a single DCI, SEP, and CCC, such that they could consider the items to assess components of multiple PEs. In these cases, panelists were split on which PE they coded for an item cluster. When they adjudicated, panelists sometimes decided that a different PE was a better choice than the one they initially selected. However, there were multiple dependencies in panelists' codings related to the selection of PE. For example, the appropriateness of the inferences in the scoring assertions were dependent on what was being measured. Hence, if a panelist decided to change the PE, it could mean that after 30 minutes spent evaluating an item cluster, the panelist then needed to re-do the entire evaluation in light of the other PE that they determined to be more appropriate.

It is important to note while a PE highlights a single DCI, SEP, and CCC, it is likely that other DCIs, SEPs, and CCCs also interweave and contribute to achievement of the overall performance expectation. Therefore, while the PE (including a single specific element from each of the three dimensions) may be appropriately defined as the unit of analysis for assessment, the standards explicitly recognize that the elements of the three dimensions of a particular PE do not exist in isolation of other elements. For example, Appendix F of the NGSS notes in bold font that "The eight [science and engineering] practices are not separate; they intentionally overlap and interconnect." The introduction to the NGSS notes that "due to the nature of some of the practices, they could not usually be used as a stand-alone practice." An example is offered, noting that the "Asking Questions and Defining Problems" practice very reasonably occurs in the context of an investigation (Planning and Carrying Out Investigations), and could involve Analyzing and Interpreting Data, which could relate to Engaging in Argument from Evidence and/or Constructing Explanations and Designing Solutions, etcetera. Similarly, CCCs are not expected to function in isolation of one another, and their "interconnections" are referenced and highlighted in Appendix G of the standards (NGSS, 2013). As such, an assessment task that authentically addresses the NGSS PEs should not be expected to require use of only the three specific dimensions of the PE in exclusion of all others. In the context of a multi-part item cluster that presents a phenomenon with text, tables, graphs, diagrams, and other visuals, and an average of approximately 10 student interactions, it is reasonable, appropriate, and consistent with the expectations of the standards and the vision of science presented by the standards that the item cluster would invoke more than just one particular DCI, SEP, and CCC. Consequently, it is reasonable that panelists might judge an item cluster to

reasonably address more than one PE – because the item cluster requires students to engage with additional elements of the standards that go beyond the three specific dimensions within a single PE. For example, to make sense of a phenomenon related to MS-LS1 (From Molecules to Organisms: Structures and Processes) a student might reasonably need to invoke aspects of the CCCs of Patterns, Cause and Effect, and Scale, Proportion, and Quantity along with the SEPs of Developing and Using Models, Planning and Carrying Out Investigations, and Engaging in Argument from Evidence in the context of DCIs about cellular structure and function. These dimensions are included in the PEs of MS-LS1-1, MS-LS1-2, and MS-LS1-3. Grounded in the assessment purpose and design, panelists were instructed to assign a single PE for an item or item cluster. It could be reasonable, therefore, to have some panelists select different PEs, with the rationale that students were, indeed, invoking the dimensions specified within these multiple PEs. In one instance in the trial run, after adjudication, panelists noted that while one PE seemed to be more closely connected to the phenomenon presented, and while students did use the dimensions of that PE to some extent in their response, another PE was more closely connected to the work that students completed in the interactions. Because of the time demands of the coding process and the dependencies of multiple evaluative points on the selection of PE, it was decided that a panel needed to settle on the targeted PE before moving on to the other evaluative considerations. Additionally, in the context of a shared state item bank used to fulfill a variety of assessment blueprints via linear-on-the-fly (LOFT) delivery, the accuracy and appropriateness of the internal PE coding is critical. As such, it is necessary not just to have reasonable panelist agreement on a specific PE for each item/cluster but also critically necessary to have reasonable agreement with the existing internal coding. Agreement with internal coding is necessary not only to verify that each assessment task reasonably satisfies what the metadata identify as the measurement target but also to allow for appropriate evaluation of the resulting inferences about student knowledge, skills, and abilities codified in the scoring assertions.

With these considerations in mind, one coding method considered was to have panelists work item-by-item and independently assign the PE, adjudicate, then cross-check the internally coded PE. This approach was determined to not be feasible based on time constraints. Panelists work at different paces, and requiring all panelists to work in parallel, item-by-item, would rush some panelists and leave other panelists waiting for periods of time. Thus, it was decided that panelists would work on their own to determine what an item or item cluster was measuring and select a PE. Then, they were to open an on-screen tab to view the internally coded PE. If they selected the same PE, they would code the PE in the updated Web Alignment Tool (WATv2) and continue. (The WATv2 is a web-based data collection tool connected to the server at the Wisconsin Center for Education Research (WCER) at the University of Wisconsin-Madison.) If

panelists selected another PE, they were to consider the appropriateness of the internally selected PE. If they did not think that the internally coded PE was appropriate, they were to discontinue coding for that item or item cluster. It was considered a necessary condition that an item or item cluster reasonably address a particular PE. Only in that context could panelists consider if the phenomenon was appropriate, decide if any parts of DCI element(s) were missing, and evaluate the scoring assertions.

Several other more nuanced modifications were also made to promote efficiency in coding and provide additional opportunities for panelists to calibrate their thinking and enhance their understanding of relevant topics. For example, in the trial run, panelists were instructed to use their preferred method for accessing the full text of the standards: via smartphone App, NGSS website, or NSTA website interface. The intent was to allow for panelists to use whichever interface they were most comfortable using. However, use of different interfaces caused unanticipated communication confusion between and among panelists. It was decided that all panelists would be expected to use the NGSS website only (nextgenscience.org) to access the full text of the standards. A small reference card with abbreviated instructions was added to the provided materials based on panelists' requests for something they could prop up on their laptop to help keep track of protocol details such as the sequence of evaluative steps and format(s) for data recording. Panelist feedback as well as questions and responses during the trial run also helped inform aspects of the training that needed clarification and/or additional examples and information. Refinements were designed to maximize efficiency without sacrificing quality or thoroughness.

The final units of analysis were determined to be the PE (with consideration of foundation boxes and NGSS Appendices), the stand-alone item (holistically, even if it required multiple student interactions), the item cluster (holistically, even if it required multiple student interactions), and the scoring assertions (holistically). Recording data at these levels reflects the study design: items and item clusters are intended, holistically, to target a PE and scoring assertions are intended to apply to the overall item in the context of the corresponding PE. Each of these units of analysis require evaluation of subcomponent parts: the specific dimensions of the PE, the individual student interactions within an item or item cluster, and the individual scoring assertions.

Panelists

To ensure appropriate representation by science discipline, and to accommodate broad state participation, 10 to 12 educators were split into two panels per grade band. Panelist selection was expected to ensure adequate content expertise across science disciplines as well as state representation as shown in **Table 3** on the following page. Five- or six-person panels allowed for two panelists with expertise per reporting category (Life, Physical, and Earth and Space Science) represented within the item bank, with some panelists having expertise in more than one science discipline. Inclusion of engineering expertise was expected to be considered as well. Per state specifications, state officials were responsible for recruiting qualified expert educators with content expertise as potential panelists for the in-person alignment institute. State officials were advised to coordinate in their recruitment to ensure consideration for sufficient representation of the diverse populations across states, including race/ethnicity, socioeconomic, and regional factors (urban/suburban/rural). Information about participating panelists is provided in **Appendix G**. WebbAlign brought an experienced group leader to facilitate each panel.

All panelists were expected to have the following qualifications:

- thorough knowledge of their discipline (be subject matter experts)
- thorough knowledge of the NGSS as interpreted in their state
- experience with NGSS assessment considerations
- experience in the appropriate grade band science education based on NGSS
- experience with the CAI test design or willingness to review the test design and released item samples in advance of the in-person work
- willingness to express professional opinions, and listen to the professional opinions of others; willingness to agree, disagree, persuade, and be persuaded; maintain collegial, respectful, and positive professional environment

In advance of the in-person alignment institute, CAI, the state Working Group, and the study director identified and confirmed the reporting categories used for each state and assessment and confirmed all units of analysis for the alignment study. This was done via conference calls, online meetings, and email exchanges between the study director and state officials as needed. The Working Group was asked to review all data collection tools and discuss any modifications that were warranted; all materials were approved for use.

WebbAlign/WCEPS was provided access to assessment items in advance of and for use during the alignment institute. Any items/clusters that repeated across sampled test events were identified. Prior to the institute, a group was registered on the WATv2 (online data collection tool) for each of the panels. Each panel was assigned one or more group identification numbers and the group leader

was designated. Then the reporting categories and PEs were entered into the WATv2 along with the information for each assessment, including the number of items and item clusters and the weight (point value) given to each. A sequential account of the alignment study procedures is provided below.

Training and Coding

Large Group Training: Appropriate training of the panelists at an alignment institute is critical to the success of the project. A necessary outcome of training is for panelists to have a common, calibrated understanding of the Categories of Engagement language system for describing levels of complexity, a shared understanding of the structure, including dimensionality, of the standards, and a shared understanding of the coding processes and associated evaluative steps.

During the morning of the first day of the alignment institute, panelists received an overview of the structure of the assessment items and item clusters, the purpose of their work, the coding processes, the use of online interfaces to view items and record data, and general training on the science Categories of Engagement definitions used to describe content complexity. The general training at the alignment institute was crafted to contextualize the origins of DOK/Categories of Engagement (to inform alignment studies of standards and assessments) and purpose (to differentiate between and among degrees of complexity), and to highlight common misinterpretations and misconceptions to help panelists better understand and, therefore, consistently apply the language system.

In advance of the study, panelists were provided with the definitions of the four Categories of Engagement for science to read in preparation for their work. Through interactive and participatory training on-site, panelists reviewed the definitions and worked toward a common understanding of the difference between and among each of the levels of complexity. Training was designed with consideration of core tenets of contemporary learning theory, including recognition of the critical importance of engaging prior understandings, as people's existing ideas greatly influence how they make sense of new ideas and construct knowledge (Posner, et al. 1982; NRC, 2000; NRC, 2005). As such, activities elicited panelists' ideas and presented opportunities for panelists to grapple with these existing ideas as well as consider if and how their existing ideas fit with (possibly new) ideas presented. Background information was shared about the overall conceptual model of complexity and the epistemology in which the model is grounded. Through facilitated activities, panelists also worked to differentiate concepts such as content (cognitive) complexity, difficulty, and multidimensionality as well as idea of sophistication of performance and the nature of learning progressions. As part of the training, panelists practiced assigning Categories of Engagement to sample PEs and assessment tasks that were selected to foster important discussions that promote improved conceptual

understanding of the tool. Explicit clarification was provided related to potential misinterpretations of the tool to evaluate complexity, informed by experiences in the Trial Run Study as well as work with classroom teachers. Panelists' responses from study evaluation forms suggest training was effective. Panelists were asked to rate (on a 1-10 scale) how comfortable they felt about the process of assigning Categories of Engagement to standards and assessment items. Twenty six out of the 28 panelists who completed the evaluation (93%) ranked their comfort with assigning Categories of Engagement as an 8 (32%), 9 (36%), or 10 (25%).

For panelists to make reliable judgements on the degree to which an item or item cluster measures student performance as relates to a particular assessment target, they must have a shared and thorough understanding of the assessment targets. Therefore, an analysis of assessment targets is a necessary component of any study that examines the degree of alignment of assessments and expectations. This need is augmented, however, in the context of the NGSS, for which there is recognition of a lack of consensus for referents as pertains to alignment analyses (e.g. Fulmer, et al, 2018). In this case, individual PEs were defined as the assessment targets. Panelists calibrated their interpretation of the PEs as pertains to stakeholder expectations as well as what is (and is not) intended for a statewide summative assessment.

To calibrate their interpretation of complexity across grade-band groups, all panelists practiced sorting a subset of NGSS PEs into categories of complexity. Panelists discussed any differences in interpretation at their tables, facilitated by the group leader. Then, the large group discussed any differences in interpretation between and among tables, working to foster a shared understanding of the idea of complexity, or "depth," as well as a shared understanding of the overall intent and meaning of the PEs they discussed. Calibration on the concept of complexity is critical for alignment analysis work as evaluation of the "depth" of the standards is a central expectation of the alignment evidence required per ESSA.

Panelists also calibrated their understanding of what would be considered an appropriate manifestation of the three-dimensional PEs in the context of an on-demand summative assessment. Similarly, panelists worked to build a common understanding of other evaluative considerations, such as the expectation for students to *make sense of* a phenomenon in their work, i.e. that students were expected to figure something out rather than answer a question that simply uses a phenomenon as a context.

After the large-group work, panelists separated into grade band groupings to continue with a more focused discussion of the standards. Group leaders first facilitated more extensive introductions and set the tone for a collaborative,

respectful, professional work environment in which panelists were expected to share and adjudicate dissenting professional judgments. Then, in their grade-band panels, panelists evaluated the complexity of a subset of the Performance Expectations (PEs) from their grade band, first assigning a Category of Engagement individually and then participating in a consensus discussion. After completing coding and discussion of the subset, the panelists reviewed the previously assigned (by content experts) Categories of Engagement for the remaining PEs, flagging any that they thought needed to be reconsidered. Panelists then discussed any differences in interpretation of the complexity of a standard. The group leader restated any major differences among panelists and facilitated the panel in coming to an agreement as to what is the most appropriate Category of Engagement for each PE. The standards analysis by grade-band panels is a necessary component of the alignment study but also, importantly, fosters thorough, nuanced, and calibrated understanding of the standards by panelists. Consensus Categories of Engagement were then entered into the online data collection system, the WATv2. The consensus Category of Engagement values for all PEs are summarized in the Findings section of this report and listed in **Appendix A**.

Next, panelists started into the analysis of the sample test events and item batches. The item-level rating instructions provided to panelists is given in **Appendix G**. The two panels within each grade band worked together to calibrate their coding as they worked through several sample items and item clusters. Panelists coded these items/clusters independently and then adjudicated as a group, discussing any differences in interpretations. The group leader facilitated discussions and communicated any specific decision rules that arose. The group leader also provided instruction and clarification on appropriate coding procedures and best practices for effective recording of comments in the WATv2. Next, the first test form was coded by both panels within each grade band and inter-panel adjudication was completed if needed. This initial calibration work was conducted to promote consistency in coding both within and between the two panels for each grade band.

Panelists were instructed to work through each stand-alone item and item cluster as if they were the student. Then, they were to determine what the item measured, i.e. what students needed to know or be able to do in order to successfully respond to the question. Panelists considered whether a student's correct response to the stand-alone item or item cluster would allow for a reasonable inference about the student's proficiency as relates to one of the PEs for the grade band. For an item cluster, successful completion of the task was expected to require students to engage with the specific three dimensions identified in the corresponding PE, at minimum. (As described previously, the particular DCI, SEP, and CCC of a PE are not expected to exist in isolation of other DCIs, SEPs, and/or CCCs in the context of a task.) For a stand-alone item,

successful completion of the task was expected to require students to engage with at least two of the specific three dimensions identified in the corresponding PE (at minimum). As panelists worked, no internal metadata were visible. After independently identifying a PE that they thought the item addressed, panelists then accessed and considered the internally coded (vendor coded) PE given in the item metadata. If the internally coded PE was appropriate, they recorded it in the online data entry system. If panelists did not think that the internally coded PE was appropriate, they were to discontinue coding for that item or item cluster. It was considered a necessary condition that an item or item cluster reasonably address the internally coded PE. Only in that context could panelists consider if the phenomenon was appropriate, decide if any parts of DCI element(s) were missing, evaluate the scoring assertions, etc. The Findings section of this report provides summary information about the eight items (out of 440) that were flagged as not addressing the internally coded PE.

Panelists also worked individually to assign a Category of Engagement to the stand-alone item or item cluster. Panelists were instructed to consider the Complexity of Engagement required by each student interaction and to record the highest Category of Engagement that was included to ensure that coding captured the full scope of the complexity of the interactions within a stand-alone item or item cluster (defined as the unit of analysis). For each activity, panelists responded to two additional questions about the item or item cluster: one about dimensionality and one about use of a phenomenon. The boxed text below shows the evaluative questions, the instructions and criteria provided, and notes on the recording of responses for these two questions.

Notes Box 1a & 1b. Dimensionality: Does a correct response require the student to engage with the three dimensions specified within the PE? • If there are multiple DCI elements (bullet points) listed, check to see if they are all incorporated into the item/cluster. If any are missing, put this in the notes box as item 1a. • For stand-alone items, note if they are three-dimensional (write “3D”) or two-dimensional (write “2D”). If 2D, state which dimension is missing. Put this in the notes box as 1b.
Notes Box 2. Phenomenon: Does the stimulus meet the test development criteria provided for a “phenomenon”? Overall, does the stimulus presented meet the following criteria: • The phenomenon is based on a specific real-world scenario and focused enough to require students' application of a SEP in the context of a DCI and CCC as intended by the PE in order to make sense of the phenomenon. • Is grade appropriate context and complexity • Is presented in way(s) that all students can access and comprehend based on information provided (including text, graphics, data, images, animations, etc.) • Is free of cultural bias, insensitivity or depiction of unsafe situation • Is puzzling and/or intriguing for students to engage in; focused on real-world observations that students can connect with or have direct experience with • Record yes/no in the notes box as item 2. If no, explain why.

If panelists coded a stand-alone item to a PE, it indicated that the item required students to engage with at least two of the three dimensions of the PE. If panelists coded an item cluster to a PE, it indicated that the item required students to engage with all three dimensions of the PE. However, additional consideration of dimensionality was requested (Notes Box 1a & 1b as shown above) in order to capture as much information as possible, including instances in which one or more DCI elements were not evident within an item/item cluster and what dimensionality (2D vs 3D) was evident in the stand-alone items. While individual instances of these occurrences are not considered alignment issues, this design helped ensure that any trends in omissions across the item bank could be captured. Panelists entered their quantitative and qualitative codings into the revised version of the Web Alignment Tool (WATv2) data collection system. No secure test information was entered.

All panelists practiced the evaluative steps with guidance from the group leaders, were provided full narrative descriptions of each evaluative prompt and protocol for response format, and were provided a smaller printed card with abbreviated coding prompts to use as a quick reference as they became more familiar with the processes.

Each panelist was given a folder containing hard copies of the following materials:

- Study schedule/agenda with logistical information
- Screenshots and instructions for logging into the WATv2 (online data collection system)
- Screenshots and instructions for navigating and viewing items via the AIR Content Rater (item access system)
- Printed sets of grade-band PEs
- Full narrative coding instruction packet
- Quick-reference coding card
- Category of Engagement Primer with background information
- Descriptions of the Categories of Engagement for science and supporting materials
- Printed versions of the full text of the NGSS were available for panelists who preferred print vs online access.

Additionally, panelists accessed materials online:

- Panelists used the NGSS website to reference the full text of the NGSS, including the foundation boxes for each PE as well as the Appendices as needed.
- An online evaluation form for panelists to provide feedback on the study process and execution was provided via email upon completion of the in-person work.

Each of three grade band panels (elementary, middle, and high school) analyzed at least one sample test event for each state for which test events were available (CT, NH, and OR). The remaining operational items were organized into batches for panels to review. Panelists analyzed the full set of operational items and clusters available at the time of the study (**Table 3**) to allow for the most direct inferences and conclusions to be made about alignment.

Table 3. Number of Items Reviewed and Included in this Report by Grade Band for Shared Science Assessment Item Bank, July 2019

Panel	States/Panelist Make-Up	Total # of Items* (Physical, Life, and Earth and Space Domains)
Elementary (3-5)	CT, HI, ID, NH, OR, RI, VT, WV, WY	136
Middle School (6-8)	CT, HI, ID, NH, OR, RI, UT, VT, WV, WY	177
High School (9-12)	CT, HI, ID, NH, OR RI, VT, WY	127

*Here and throughout the tables within this report, “items” refers to both stand-alone items and item clusters. While both may require multiple student interactions, the unit of analysis is the overall item or item cluster.

Panelists were instructed to focus primarily on the alignment between the PEs and the assessment items and item clusters provided. However, panelists were able to provide qualitative input or feedback on the PEs and on the assessment items and clusters by writing a note in the appropriate text box in the WATv2 data collection tool. Panelists entered commendations, recommendations, and extensive editorial feedback. Panelists could indicate whether there was a Source of Challenge issue with an item—i.e. a technical or content problem with the item that might cause the student who knows the material to give a wrong answer or enable someone who does not have the knowledge being tested to answer the item correctly. After a panelist completes coding all of the assessment items and item clusters within a batch, the WATv2 offers a set of debriefing questions to answer for each study. These questions solicit feedback from the panelists about the assessment as a whole, and about topics that are not captured in the item-level coding data.

If needed and as time allowed, the results for each study were adjudicated after all of the panelists completed coding a test form or batch of items. The adjudication process helped to ensure that the coding by panelists did not include spurious data and that the codes entered were those as intended. For example, adjudication can correct errors, such as if a panelist accidentally entered one standard but meant to enter another standard. Group leaders facilitated conversations about items or item clusters for which panelists differed

significantly on the Category of Engagement assigned (e.g. three different values assigned). When these substantial differences in coding occur, it sometimes indicates a data entry error. If data are entered as intended, then it suggests that panelists are either interpreting the Category of Engagement definitions in very different ways or are interpreting the particular assessment item in very different ways. For PE assignment, only data entry issues were addressed in final adjudication as any clarifications or discussion of differences in perspective on PE selection were addressed as needed as panelists moved through a test form or item batch. Panelists did not conduct adjudication specific to the evaluative prompts for Use of Phenomenon or for the evaluations of the relationships with Scoring Assertions, but sometimes discussed these codings in the context of overall discussion about an item. Overall, adjudication was conducted to foster full and appropriate interpretation of the assessment items/clusters and to ensure that panelists had coded the items/clusters as they intended. Very limited adjudication occurred after the first couple of batches for each panel. Panelists were not required to change their coding after the discussions. Panelist agreement statistics were computed after adjudication and are included in the Findings section of this report.

The study director and group leaders monitored the data entry process by having access to the administrator section of the WATv2. This monitoring included noting the progress each panelist was making and if there were any irregularities in the data entry where an intervention was needed. Examples of problems that could cause irregularities and, therefore, require intervention, include human error, such as a panelist accidentally skipping an item and then entering data out of sequence, as well as misunderstanding of the processes. At the end of each day, the group leaders met with the study director and state representatives to discuss the progress in coding made during the day and to discuss any questions that may have arisen.

Data Analysis

The eight items (out of 440) that were not considered to reasonably target the internally coded PE are tabulated in the Findings section of this report and identified individually in **Appendix C**. To derive the results from the analysis of item complexity, the panelists' codings for Categories of Engagement were averaged. First, the value was computed for each individual panelist. Then the final reported value for each criterion was found by averaging the values across all panelists. Any variance among panelists was considered legitimate, for example, with the reported Category of Engagement for an item falling somewhere between the two or more assigned values. Such variation could signify differences in interpretation of an item or of the assessed content and/or a Category of Engagement that falls in between two of the four defined levels. Any large variations among panelists in the final results represented true differences in opinion among the panelists and were not because of coding error. Standard

deviations are reported in the tables provided in **Appendix B**, which give one indication of the variance among panelists. Majority coding (at least 3 out of 5 or 4 out of 6 panelists) was used to determine whether the criteria of Dimensionality and Use of Phenomena were met as well as for the two evaluative questions related to the scoring assertions.

The results from this study pertain specifically to the issue of alignment between the NGSS PEs and the assessment sample test forms and items/clusters that were analyzed. An alignment analysis of this nature does not serve as external verification of the general quality of the standards or assessments. While some feedback is provided on aspects of quality, the degree of alignment is the focus of the discussion in the results. It is important to note that the design of the items and item clusters in this program include elaborate stimuli made up of narratives, tables, graphs, and other visuals, extensive student interactions that include additional text tables, graphs, and other visuals, and detailed narrative scoring assertions. The scope of material incorporated into each item and item cluster means abundant opportunities for typos, confusing wording, layout issues, mislabeling, and other editorial issues. Panelists commented on a multitude of corrections and other editorial issues that have implications for student responses and scoring and therefore require or should be considered for revisions.

Alignment Criteria Used for This Analysis

When using the traditional Webb methodology for analysis of the alignment of curriculum standards with on-demand summative assessments for purposes of submitting evidence of alignment for federal peer review, results have typically been reported according to four alignment criteria (Categorical Concurrence, Depth of Knowledge Consistency, Range of Knowledge Correspondence, and Balance of Representation). Specific cutoffs were typically assigned to each criterion using defined decision rules about what was considered acceptable. The rationales for all decision rules were provided to allow for a process that was as transparent as possible. States were always asked to modify these levels of acceptable alignment as warranted.

In the context of NGSS alignment, the same four criteria (Categorical Concurrence, Depth of Knowledge Consistency, Range of Knowledge Correspondence, and Balance of Representation)—considered as broad categories of types of information—still are relevant in terms of the general types of information necessary to judge degree of alignment and as required for submission to federal peer review. However, adjustments need to be made to the specifics of each criterion to appropriately correspond to NGSS and other three-dimensional and Framework-based standards. For example, it is still necessary to consider if an assessment concurs with the categories of expectations in the standards, but it is also necessary to consider dimensionality as a requirement of that concurrence. Similarly, it is still expected that a test assesses some range, or breadth, of the standards, and that there is some balance of standards' representation—but these criteria must be reframed in terms of the NGSS context as well as in terms of the particular assessment blueprint and construct. For example, appropriate range would be expected to manifest differently on an assessment designed to assess Life Science PEs *only* compared with on an assessment designed to assess PEs across *all* domains.

Additional criteria must be also be included that correspond to the specific intents and claims of an NGSS statewide summative assessment. In the context of the NGSS, for example, the structure of how students are to know, engage, and think about science is very relevant to how the measurement of knowledge should be designed. The degree to which the three-dimensional engagement as expressed in the standards is captured on the assessment is a question of structure and is reflected in the Dimensionality/Structure of Knowledge Comparability criterion. A content analysis is also required to provide evidence of the Use of Phenomena as well as the relationship of scoring assertions to the PE and to the student interactions, relevant to the specifics of the assessment design.

New decision rules and cutoffs for acceptable alignment also must be defined. For example, the cutoff that has typically been considered acceptable for breadth (Range of Knowledge Correspondence) in the context of an on-demand summative assessment test form does not have as much meaning if used with NGSS because of the vastness and structure of the standards. In the context of NGSS and other Framework-based standards, it is more important to confirm that the assessment in some way measures student performance as intended by the standards and within the intentions of a state's assessment design. For states sharing this item bank, breadth needs to be considered from the perspective of an individual student (by test form) as well as for an overall student population (by aggregate data from all administered assessments).

After input from and discussion with representatives from all states, the criteria detailed in **Table 4** were agreed upon to be used to report on the degree of alignment of standards, scoring assertions, and the actual assessment items/clusters. The identification of the set of criteria was grounded in discussions with states to clarify the full set of claims that states intended to make as relates to the assessments that would require content alignment analysis. For example,

- All states expected the assessment items/clusters to be phenomenon-based, grounded in a shared definition of this condition.
- All states expected that item clusters are three-dimensional and stand-alone items are two-dimensional or three-dimensional.
- All states expected the assessment items to reasonably target the specific elements within the foundation boxes for each dimension of a particular PE.

In general, acceptable levels for breadth and depth as used in the state-specific analyses are anchored in each state's assessment claims. Each state was asked to review the proposed decision rules for determination of acceptable cutoffs for each alignment criterion and approve, or request modifications if warranted. Acceptable alignment cutoffs may vary by state. For example, factors that will affect what is considered appropriate breadth include the test construct (e.g. one domain (Life Science) for Hawai'i's high school test vs three domains (Life, Physical, and Earth and Space Science) for other states' high school test), and other relevant blueprint specifications. In the case of criteria which states expect to be met for **all** items/clusters, a 90% cutoff is used to allow some leeway for human error and differences in professional opinion.

Table 4. Consensus Alignment Criteria for Shared Science Assessment Item Bank, 2019

Criterion	Intended Claim/Inference	Acceptable Cutoff
Use of Phenomena	Items/clusters require students to engage multiple dimensions of the standards (“use science”) to make sense of phenomena. Each assessment item/item cluster is grounded in a stimulus that meets the test development criteria for a phenomenon.	At least 90% of items/clusters are considered phenomenon-based by a majority of panelists (i.e. at least 3 out of 5 or 4 out of 6 panelists).
Dimensionality/ Structure of Knowledge Comparability	Item clusters require students to demonstrate integrated engagement with the three dimensions of SEPs, DCIs, and CCCs specified in the targeted PE. Stand-alone items require students to demonstrate integrated engagement with two or three of the dimensions specified in the targeted PE.	At least 90% of clusters are considered three-dimensional by a majority of panelists; at least 90% of stand-alone items are considered multi-dimensional by a majority of panelists.
Categorical Concurrence*	Test events have the potential to yield sufficient evidence to make inferences about student knowledge, skills, and abilities (KSAs) as relates to each reporting category.	A test form will include at least six (6) opportunities to respond to items or clusters that target the standards within each reporting category.
Consistency of Cognitive Engagement	The assessment elicits work that is as cognitively demanding as the expectations in the standards.	While some individual student interactions may be Category 1, no items/clusters should include only Category 1 interactions. Proportions of items/clusters with Category 2 and 3 opportunities reflect grade band PEs. Some aspects of Category 4 PEs will be assessed but the full scope of Category 4 PEs is expected to be assessed in the classroom.
Range of Knowledge Correspondence (Population)*	State-specific claims will be considered against aggregate data from all administered test events in the state in conjunction with external confirmatory analysis of vendor metadata. Confidence in vendor metadata depends on results of external confirmatory analysis.	At least 90% of PEs have the potential to be assessed across the student population.

Table 4 Cont'd. Consensus Alignment Criteria for Shared Science Assessment Item Bank, 2019

Criterion	Intended Claim/Inference	Acceptable Cutoff
Range of Knowledge Correspondence (Individual)*	Test events assess an appropriate breadth of the standards, as defined by the state assessment blueprint. Assessed standards are sampled across topics within each reporting category for individual students.	Test forms analyzed meet blueprint specifications for Range. Blueprints are expected to specify sampling across topics (or other sublevels for each reporting category).
Balance of Representation*	No PE is targeted more than once on any test event.	A PE should not be targeted more than once on a test event; each stand-alone item and item cluster should target a different PE.
Relationship of Scoring Assertions with Student Interactions	In aggregate, the scoring assertions for an item/item cluster appropriately represent the inferences about student knowledge, skills, and abilities that can be made based on successful interactions with an item/cluster.	For at least 90% of all items/clusters, a majority of panelists consider a large majority of the scoring assertions (at least ~75%) to appropriately represent the inferences about student KSAs that can be made based on successful interactions with an item/cluster.
Relationship of Scoring Assertions with PEs	In aggregate, the scoring assertions for an item/item cluster appropriately represent the three-dimensional expectations of the targeted PE. (At least two of the three dimensions for stand-alone items.)	For at least 90% of all items/clusters, a majority of panelists consider a large majority of the scoring assertions (at least ~75%) to appropriately represent the expectations within the corresponding PE.
Blueprint Analysis*	Test blueprints demonstrate the capacity to yield aligned test events.	Results are described; any potential alignment gaps identified.

*State-specific results for these criteria are included in the state-specific reports.

No states made specific claims about:

- Engineering Design (ETS) standards or engineering (Engineering practices are incorporated in two to nine PEs per grade/grade band. These standards were included in the study. The ETS standards were not included in the study.)
- Science, Technology, Society, and the Environment Connections
- Nature of Science

Details on the criteria used for determining the degree of alignment between standards and assessments are provided below. For each criterion, the cutoff for acceptability is defined. If results meet these defined cutoffs, the criterion is considered to be met. Typically, a criterion is considered to be “weakly met” if results fall within 10% of the expected cutoff.

Reporting Categories and Standards

Study results are reported using each domain as a reporting category (RC). For most states and assessments, the reporting categories are Physical Sciences, Life Sciences, and Earth and Space Sciences. Consensus Category of Engagement values for all PEs are given in **Appendix A** for each subject. All PEs for the domains of Physical Sciences, Life Sciences, and Earth and Space Sciences were included in the analysis. In the descriptions below, the term “standards” may be used as an umbrella term, to refer to expectations in general.

Use of Phenomena

Assessments that draw from the Shared Science Assessment Item Bank are intended to be phenomenon-based, meaning that items/clusters require students to engage multiple dimensions of the standards (“use science”) to make sense of phenomena. The test design specified a set of criteria that defined an appropriate phenomenon, including that it is based on a specific real-world scenario, reflects grade-appropriate content and complexity, and is focused enough to require students' application of a SEP in the context of a DCI and CCC as intended by the PE in order to make sense of the phenomenon (for an item cluster). While stand-alone items may be two *or* three dimensional, they were still expected to require students to use multiple dimensions of a PE to make sense of a phenomenon. To meet this expectation at an item level, a majority of the reviewers on a panel (at least 3 out of 5 or 4 out of 6) must have considered the item or item cluster to have met the test development criteria for a phenomenon, as indicated in their independent coding. To meet this criterion overall, at least **90%** of items/clusters must have been considered phenomenon-based by a majority of panelists. Although states expect this criterion to be met for ***all*** items/clusters, a 90% cutoff is used to allow some leeway for human error and differences in professional opinion.

Dimensionality / Structure of Knowledge Comparability

All assessment items in the Shared Science Assessment Item Bank are intended to be two- or three-dimensional, meaning that items/clusters require a student to engage with and interweave two or three dimensions of the NGSS (Disciplinary Core Ideas, Science and Engineering Practices, and Crosscutting Concepts) as the student makes sense of phenomena. Item clusters are expected to require students to demonstrate integrated engagement with, at minimum, the specific three dimensions specified in the targeted PE. Stand-alone items are expected to require students to demonstrate integrated engagement with, at minimum, two or three of the dimensions specified in the targeted PE. To meet this expectation at an item level, a majority of the reviewers on a panel (3 out of 5 or 4 out of 6) must have indicated in their independent coding that an item cluster required student engagement with all three dimensions of the PE or that a stand-alone item required student engagement with two or three of the dimensions. To meet this criterion overall, at least **90%** of items/clusters must have been considered to

have the intended dimensionality by a majority of panelists. Although states expect this criterion to be met for **all** items/clusters, a 90% cutoff is used to allow some leeway for human error and differences in professional opinion.

Categorical Concurrence

The NGSS are organized by content categories including Physical, Life, and Earth and Space Sciences. Each of these categories is further divided by content (DCIs). An important aspect of alignment between standards and assessments is whether both address the same content categories. The Categorical Concurrence criterion provides a very general indication of alignment if both documents incorporate the same content. The criterion of Categorical Concurrence between standards and assessments is met if the same or consistent categories of content appear in both documents. For a particular assessment form or test event, this criterion would be judged by determining whether the assessment included items targeting PEs from each reporting category. Grounded in calculations based on a procedure developed by Subkoviak (1988), it is typically assumed that an assessment would have to have at least six items for measuring content from a reporting category for a minimum acceptable level of Categorical Concurrence to exist between the domain and the assessment (Webb, 1999). The number of items (six) is based on estimating the number of items that could produce a reasonably reliable score for estimating students' mastery of content on that subscale. Of course, many factors must be considered in determining what a reasonable number is, including the reliability of the subscale, the mean score, and cutoff score for determining mastery. A cutoff of six items per reporting category was consistent with state expectations and was used in this analysis. This criterion is reported on primarily in state-specific reports. In this common report, Categorical Concurrence of the grade band item banks by domain is summarized in the Findings.

Consistency of Cognitive Engagement (Category of Engagement)

A Framework for K-12 Science Education and the resulting NGSS both emphasize a conceptual shift in science standards, related to the complexity of student engagement with science concepts and scientific thinking (NGSS Appendix A, Conceptual Shift #4). As a central conceptual shift, attention must be given to determine if and in what ways different types of student cognitive engagement (i.e. cognitive complexity) are being interpreted both in the expectations and the assessment. Consistency of Cognitive Engagement between content standards and an assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the corresponding standards. For consistency to exist between the assessment and the reporting categories, as judged in this analysis, two conditions apply. First, no items or item clusters should require only Category 1 Cognitive Engagement. Category 1 expectations and tasks require recall of facts and terms, recognition of structures and

properties, reproduction of standard scientific representations, the performance of routine procedures (e.g. using measurement tools), and other rote or routine work. While it is acceptable for some interactions within an item or item cluster to be Category 1, successful completion of an item or item cluster cannot require only Category 1 work, per state expectations. This requirement is consistent with the intent of the NGSS, which emphasize that the standards represent “...a huge transition, from a focus on knowledge itself to a focus on putting...knowledge to use—a transition that in and of itself necessitates a corresponding leap in rigor” and notes that new standards “focus on understanding rather than memorization” (NGSS Appendix C, 2013). While the standards recognize that Category 1 work is necessary in the classroom (NGSS Appendix B), an explicit goal of Framework-based standards, including NGSS, is to promote a shift away from assessing students on Category 1 types of tasks. Second, the proportion of items and item clusters with Category 2 and Category 3 opportunities should reflect the proportion of Category 2 and Category 3 expectations in the Performance Expectations. Category 4 expectations, which are complex tasks that require extended time (such as the “sustained investigations” expected by the Framework) are not expected to be appropriately or authentically assessed in an on-demand context. All of the items and item clusters in the Shared Science Assessment Item Bank, therefore, are expected to provide opportunities for cognitive engagement within Category 2 and Category 3. Although this is expected for **all** items/clusters, a 90% cutoff is used to allow some leeway for human error and differences in professional opinion.

This alignment expectation for the Consistency of Cognitive Engagement is grounded directly in the vision of the standards. Category 2 tasks require students to connect ideas and make sense of relationships and interactions between and among concepts and ideas, anchored in evidence-based thinking. Category 3 tasks involve abstract, analytical, hypothetical, critical, evaluative, original (to the student), and innovative thinking, including crafting reasoned scientific arguments based on evidence. The NGSS focus on the types of cognitive engagement included within Category 2 and Category 3 is emphasized in multiple places within the Framework and the resulting NGSS. For example, the Framework committee emphasized the expectation for students to “achieve depth of understanding of the core ideas” (NRC, 2012). Similarly, the standards specify a “focus on deeper understanding of content as well as application of content” and reiterate that “the NGSS focus [is] on understanding rather than memorization” as well as “putting...knowledge to use” (NGSS, 2013). These types of cognitive engagement are explicated in the Category 2 definitions for Science. Expectations for Category 3 engagement are also called out within the NGSS documentation. For example, the Framework committee specified that the standards should allow opportunities for “students to engage in scientific...argumentation” (NRC, 2012) and the standards specify the goal of supporting students as they “discove[r] new knowledge, solv[e] challenging

problems, and generat[e] innovations” including addressing “problems not previously encountered” (NGSS Appendix C, 2013). These types of cognitive engagement are explicated in the Category 3 definitions for Science.

Category of Engagement 1-4

Interpreting and assigning Category of Engagement (types of complexity) to both standards and assessment items is an essential requirement of alignment analysis. The Category of Engagement descriptions help to clarify how different types of complexity are represented in the sciences. Full descriptions for science are included in **Appendix F**.

Range of Knowledge Correspondence (Individual and Population)

Traditionally, Range of Knowledge is calculated against the full scope of a set of assessed academic standards, and a test form is expected to sample KSAs from at least half of the full set of standards. In the context of NGSS, it is not reasonable to consider the full range of grade-band standards across multiple disciplines as the referent, both because of the vast scope of the standards as well as because the standards are intended to foster deep engagement with science versus broad coverage of topics. Instead, state stakeholders define what is appropriate and reasonable for assessment on an individual test event. The intended range is then codified in the test blueprint, which serves as the referent. For reporting categories and assessments to be aligned, the breadth of knowledge expected on the test blueprint should be comparable to the breadth of knowledge sampled on a test form. In other words, the span of knowledge expected of students by a reporting category (as defined by a test blueprint) should correspond to the span of knowledge that students need to correctly answer the assessment items. Because the test blueprint serves as the referent, fidelity to blueprint specifications can serve as evidence for meeting this alignment criterion, interpreted in the context of the results of the item-level content analysis. Test blueprints are organized by Domain and by DCI, and specify the expected range of sampling within each DCI as well as across all DCIs within the Domain. In the context of this assessment program, the criterion of Range must be considered for the individual student as well as for the overall tested population. State-specific claims were considered against aggregate data from all administered test events in the state. These findings were interpreted in conjunction with the item-level analysis, which was used to determine the degree to which the claims within the item metadata could be substantiated. Range of Knowledge at the test event level (for individual students) is addressed in state-specific reports. In this common report, the overall Range of Knowledge represented by the item bank is summarized. States expected at least 90% of the PEs to have the potential to be assessed across the student population.

Balance of Representation

In addition to comparable depth and breadth of knowledge, aligned reporting categories and assessments require that knowledge be distributed in the intended proportions. The Balance of Representation criterion, as applies to the design of the assessments derived from the Shared Science Assessment Item Bank, specifies that no PE is targeted more than once on any single test event. This expectation is considered against aggregate data from all administered test events in each state.

Relationship of Scoring Assertions with Student Interactions

Each stand-alone item and item cluster is scored with a set of binary (true/false) narrative scoring assertions, which constitute the scoring rationales for items. Per CAI, each assertion describes a piece of content knowledge, skill, or ability (KSA) that is related to the targeted PE and that the student is expected to have demonstrated by successful interaction with the item. In general, an assertion states the student's action(s) within the item that provide(s) evidence for the corresponding inference about student KSAs. In aggregate, the scoring assertions for an item/item cluster should appropriately represent the inferences about student knowledge, skills, and abilities that can be made based on successful interactions with an item/cluster. To meet this expectation at an item level, a majority of the reviewers on a panel (3 out of 5 or 4 out of 6) must have indicated in their independent coding that a large majority (~75%) of the scoring assertions describe a direct inference that can be made from the student's correct response. To meet this criterion overall, at least **90%** of all items/clusters must be considered to have scoring assertions (in aggregate) that appropriately represent the inferences about student KSAs that can be made based on successful interactions with an item/cluster. Although states expect this criterion to be met for **all** items/clusters, a 90% cutoff is used to allow some leeway for human error and differences in professional opinion.

Relationship of Scoring Assertions with PEs

As scoring rationales, the assertions are expected to appropriately reflect the assessment targets (the PEs). In aggregate, the scoring assertions for an item cluster are expected to appropriately represent the three-dimensional expectations of the targeted PE. For stand-alone items, the assertions may reflect either two or three dimensions of the PE. To meet this expectation at an item level, a majority of panelists must have indicated in their independent coding that the scoring assertions, in aggregate, represented the expectations explicit within the corresponding PE. For at least **90%** of all items/item clusters, the scoring assertions (in aggregate) are considered to appropriately represent the expectations within the corresponding PE. Although states expect this criterion to be met for **all** items/clusters, a 90% cutoff is used to allow some leeway for human error and differences in professional opinion.

Source of Challenge Criterion

The Source of Challenge criterion is used to identify items on which the major cognitive demand is inadvertently placed and is other than the targeted reporting category or expectation (i.e. construct irrelevance). Bias and sensitivity issues, as well as technical issues and errors, could all be reasons for an item to have a Source of Challenge problem. Such item characteristics may result in some students not answering an assessment item or answering an assessment item incorrectly even though they possess the understanding and skills being assessed.

Summary Findings: Shared Science Assessment Item Bank

The results and a discussion of both the standards analyses and the item-level analyses are presented in this section. In order for student scores on an assessment to support the intended inferences about student achievement (scoring assertions) as relates to the NGSS, there must be a close underlying relationship between and among the standards, the assessment items, and the scoring assertions. The results of the standards analyses are reported first, to provide context that can support interpretation of the results of the item-level analysis. Then, the results of the item-level analysis of the Shared Science Assessment Item Bank are reported by grade band.

Overall, the item-level analysis found that the shared item bank for each grade band showed strong capacity to generate aligned test events (pending state-specific documentation), as summarized in **Table 5** below.

Table 5. Overall results by alignment criterion and grade band for Shared Science Assessment Item Bank, 2019

Criterion	Is the criterion met for each grade band?		
	Elementary	Middle	High
Use of Phenomenon	YES	YES	YES
Dimensionality/Structure of Knowledge	YES	YES	YES
Categorical Concurrence*	YES	YES	YES
Consistency of Cognitive Engagement*	YES	YES	YES
Range of Knowledge (Population)*	YES	YES	WEAKLY**
Balance of Representation*	YES	YES	YES
Relationship of Scoring Assertions with Student Interactions	YES	YES	YES
Relationship of Scoring Assertions with PEs	YES	YES	YES

*Item bank indicates potential to meet these criteria but state-specific evidence and final determinations by state are given in the state-specific reports; Range of Knowledge for individual test forms is given in state-specific reports.

**For the high school Physical Science domain, this criterion was unmet. If considering the full set of high school standards, this criterion was weakly met.

In general, full or acceptable alignment was found between the NGSS PEs and the assessment items and item clusters for all grade bands although specific items (two item clusters and 18 stand-alone items) were identified that did not meet one or more alignment expectations, and warrant revisions or removal. Even for items that met alignment expectations, many editorial suggestions were made to correct errors in text and graphics, clarify content or instructions, and/or address scientific inaccuracy. Overall, however, panelists found that items met states' expectations that assessment tasks required integrated engagement with at least two (stand-alone items) or three (item clusters) dimensions of SEPs, DCIs, and CCCs specified in the targeted PE in order to make sense of a phenomenon, and engaged students at appropriate levels of complexity.

The vast majority of PEs were represented by items and/or item clusters within the overall item bank. The elementary grades item bank contained items addressing all of the Grades 3-5 PEs. However, the one item targeting 4-ESS3-1 was flagged by panelists for issues. Even omitting this item, 98% of the elementary grades PEs were found to have between one and six corresponding items. The middle school item bank contained items addressing 96% of the Grades 6-8 PEs, even when taking into account any items flagged by panelists. Each of these PEs was represented by between one and eight items. The high school item bank contained items addressing 84% of the Grades 9-12 PEs, again, taking into account the one high school item flagged by panelists. Each targeted PE was represented by between one and six items. The Physical Science domain was underrepresented in the high school item bank, with items targeting only 66% of the PEs. Otherwise, items corresponded to PEs across all domains, with no domain over or underemphasized. Because panelists found between one and eight items corresponding to each targeted PEs, no PE was considered overemphasized in the item bank.

Results suggest the high school item bank needs to be supplemented if states expect item banks to have the capacity to assess at least 90% of the standards. This issue could be fully resolved with the addition of at least six items to the high school item bank. At least five of these items would need to address unrepresented PEs within the Physical Science domain.

Standards

A summary of the Categories of Engagement of the NGSS PEs is given in **Table 6** by grade band. All PEs were judged to have a complexity Category of 2, 3, or 4. Across grade bands, the vast majority of PEs were considered Category 2 (71% to 76%). Between 15% and 24% of the PEs for each grade band were considered Category 3. The remaining PEs were considered Category 4 (5% in elementary and middle school; 15% in high school). An explanation of the types of expectations encompassed by each Category of Engagement, along with an explanation of why the absence of Category 1 PEs is consistent with the NGSS intent, is given in the following paragraphs.

The Categories of Engagement tool is used to differentiate between and among the different types of complexity of cognitive engagement required by learning expectations and tasks. Category 1 includes tasks such as recalling facts and terms, recognizing structures or properties, reproducing standard scientific representations, or performing routine procedures. The Framework and NGSS documentation specify that Category 1 type expectations are not intended as assessment targets. For example, NGSS Appendix C calls out “...a huge transition, from a focus on knowledge itself to a focus on putting...knowledge to use—a transition that in and of itself necessitates a corresponding leap in rigor” and notes that new standards “focus on understanding rather than

memorization.” Because “[p]erformance expectations are the assessable statements of what students should know and be able to do” and are intended “to make clear the intent of the assessments” (NGSS, 2013) it can be inferred that no PE should be considered to expect only Category 1 type work. Although students should not be assessed on Category 1 tasks, the NGSS recognizes that students will indeed need to engage with Category 1 tasks in the context of broader work to make sense of a phenomenon. For example, although “[n]o part of the NGSS specifies the student outcome of defining a gene – it is...implicit that in order to demonstrate proficiency on MS-LS3-1, students will have to be introduced to the concept of a gene through curriculum and instruction” (NGSS Appendix B, 2013). Similarly, students will need to use particular measurement tools, recall appropriate safety protocols, and learn new terms. Overall, students may be expected to develop fluency with Category 1 expectations but they are not appropriate as overall summative assessment targets. Because of this clear expectation within the standards, it is critical that educators and assessment developers can consistently differentiate between Category 1 and Category 2 tasks.

Category 2 tasks require students to connect ideas and make sense of relationships and interactions between and among concepts and ideas, anchored in evidence-based thinking. The conceptual understanding emphasized by Category 2 expectations are reflected in multiple places in Framework and NGSS documentation. For example, Appendix A conceptual shift number four states that “[t]he NGSS focus on deeper understanding of content as well as application of content” (NGSS Appendix A, 2013). Appendix C also underscores this key shift, noting that “the NGSS focus [is] on understanding rather than memorization” (NGSS Appendix C, 2013). This, in turn, reflects the Framework committee’s intent to “give time for students to...achieve depth of understanding of the core ideas” (NRC, 2012). A core overall goal of Framework-influenced standards, including NGSS, is for students to demonstrate knowledge-in-use as they make sense of phenomena, consistent with many Category 2 types of expectations.

Category 3 tasks involve abstract, analytical, hypothetical, critical, evaluative, original (to the student), and innovative thinking, including crafting reasoned scientific arguments based on evidence. Category 3 expectations are reflected in the Framework committee’s intent to “give time for students to engage in scientific...argumentation” (NRC, 2012) and support the goal of supporting students as they “discove[r] new knowledge, solv[e] challenging problems, and generat[e] innovations” including addressing “problems not previously encountered” (NGSS Appendix C, 2013).

Category 4 tasks expect at least the complexity of Category 3 but require extended, and iterative sensemaking, corresponding to the “expectation...that students generate and interpret evidence and develop explanations of the natural

world through sustained investigations” or that students “carry out empirical investigations in order to develop or evaluate knowledge claims” (NRC, 2013). While subcomponents of Category 4 tasks may be represented in an on-demand assessment, they are more appropriately and authentically assessed in the classroom.

Consistency of Cognitive Engagement

The last two columns of **Table 6** show the distribution of standards at each Category of Complexity (2, 3 or 4) next to the distribution of assessment items (both stand-alone and item clusters) within the item bank at each Category of Complexity (1, 2, 3 or 4) by grade band. This allows for a broad-stroke look at the overall complexity of the items within the item bank for each grade-band.

Although no items or item clusters were expected to include only Category 1 interactions, a very small percentage of elementary and middle school items (for each, 4 items or 2-3%) were identified as such. While this information can be used for ongoing improvements to the item bank, it is not considered a threat to the alignment of test events with PEs. Items and item clusters were found to be Categories 2 and 3, corresponding proportionately, overall, to the complexity of the grade-band PEs.

Table 6. Performance Expectations by Category of Engagement compared with Category of Engagement distribution within grade-band item bank samples

NGSS Grade Bands	Total Number of PEs	Category of Engagement	Number of PE by Level	Standards % Category of Engagement Distribution by Level	Items % Category of Engagement Distribution by level
Grades 3-5	42	1	0	0	3
		2	32	76	77
		3	8	19	20
		4	2	5	0
Grades 6-8	28	1	0	0	2
		2	39	71	55
		3	13	24	43
		4	3	5	0
Grades 9-12	28	1	0	0	0
		2	48	72	80
		3	10	15	20
		4	9	13	0

The results of the item-level analysis suggest that the item bank has the capacity to meet the criterion of Consistency of Cognitive Engagement, which expects the assessment to elicit work that is as cognitively demanding as the expectations in the standards. Consistency of Cognitive Engagement by domain and grade band is shown in **Table 7**.

Table 7. Consistency of Cognitive Engagement by Domain and Grade Band Based on Overall Item Bank

Consistency of Cognitive Engagement by Grade and Reporting Category						
	Grades 3-5		Grades 6-8		Grades 9-12	
PS	At	Above*	At	Above	At	Above
	60%	19%	38%	42%	53%	11%
LS	At	Above	At	Above	At	Above
	73%	25%	53%	40%	66%	16%
ES	At	Above	At	Above	At	Above
	80%	13%	45%	46%	44%	31%

*This indicates that there was at least one opportunity for students to engage at a Category 3 even when an item or item cluster targeted a Category 2 PE.

As new “rigorous” standards have driven forward expectations for complexity, on-demand assessments have become more successful in addressing higher complexity expectations. The item cluster structure of this assessment, along with the multiple types of information provided within the stimulus and the multiple types of student interactions possible, provided opportunities for students to engage in a wide variety of complex tasks. Between 97% and 100% of items were considered Category 2 or 3, in excess of the 90% cutoff for acceptability of alignment. For each domain and grade band, between 11% and 47% of items/clusters were found to require Category 3 engagement for at least one or more of the student interactions while the corresponding PE was considered Category 2. At first glance, this could be interpreted to mean that the assessment is overly complex. However, if taking into account that 95% of the Category 3 ratings were for item clusters, then the “above” proportion should not be considered a concern. Across all grade bands, only four stand-alone items were coded as Category 3. The other 127 instances of Category 3 were all for item clusters. Panelists reported that typically, when assigning an item cluster a Category 3, most of the interactions within an item cluster were considered Category 2 (and sometimes also one or two Category 1 interactions) but that at least one of the interactions required students to interweave the components of the tasks such that there was opportunity to engage at Category 3. Panelists were instructed to code an item cluster (defined as the unit of analysis) to the highest Category of Engagement that it included to ensure that coding captured the full scope of the complexity of an item cluster. Because the item bank is comprised of appropriately complex items and item clusters, reflecting the

distribution of the complexity within the PEs, it suggests strong capacity for test forms to meet the Consistency of Cognitive Engagement criterion.

Use of Phenomena

Assessments that draw from the Shared Science Assessment Item Bank are intended to be phenomenon-based, meaning that items/clusters require students to engage multiple dimensions of the standards (“use science”) to make sense of phenomena. The test design specified a set of criteria that defined an appropriate phenomenon, including that it is based on a specific real-world scenario, reflects grade-appropriate content and complexity, and is focused enough to require students’ application of a SEP in the context of a DCI and CCC as intended by the PE in order to make sense of the phenomenon (for an item cluster). While stand-alone items may be two *or* three dimensional, they were still expected to require students to use multiple dimensions of a PE to make sense of a phenomenon. With the exception of several items flagged for revision or removal because all interactions were Category 1 (see **Table 11**), all items were considered by a majority of panelists to meet states’ expectations for being phenomenon-based. (If an item is coded as Category 1, it means that the item does not require a student to interact with the phenomenon presented, and therefore means that even if a specific real-world scenario is presented, it does not meet the full set of expectations for Use of Phenomena.)

Categorical Concurrence

Overall results as pertains to Categorical Concurrence of the grade band item banks by domain are summarized in **Table 8**. For each grade band and domain, the total number of items is given. All PEs with no corresponding items are listed in **Table 9**. If a PE had one corresponding item but that item was flagged by panelists because it did not meet one or more of the evaluative criteria, then the PE is listed as unrepresented. Results reported here apply across states. Relationships between state-specific PEs and items are addressed in the state’s report.

Table 8. Number of Shared Science Assessment Items Included in this Report by Domain and Grade Band Based on Item-Level Analysis of Overall Operational Item Bank

Number of Items by Grade and Reporting Category			
	Grades 3-5	Grades 6-8	Grades 9-12
PS	47	59	42
LS	45	68	52
ESS	41	50	37

Table 9. Unrepresented PEs in Overall Operational Item Bank by Domain and Grade Band Based on Item-Level Analysis of 440 Items Included in this Report

PEs Not Represented in the Item Bank			
	Grades 3-5	Grades 6-8	Grades 9-12
PS	All PEs represented	MS-PS4-3	HS-PS1-8 HS-PS2-5 HS-PS2-6 HS-PS4-2 HS-PS4-3 HS-PS4-4 HS-PS4-5
LS	All PEs represented	MS-LS4-6	HS-LS2-5 HS-LS2-7
ESS	4-ESS3-1	All PEs represented	HS-ESS2-5 HS-ESS3-4

Range of Knowledge Correspondence

The criterion of Range of Knowledge shows the breadth of standards within each reporting category that are assessed by one or more items within a test form or that are represented within an item bank or item sample. **Table 10** shows the distribution of items in the overall item bank. The total number of PEs per grade is shown next to the number of PEs not represented. The rightmost column shows the percentage of PEs with one or more corresponding items by grade band and domain. States expected at least 90% of the PEs to have the potential to be assessed across the student population.

Table 10. Shared Science Assessment Item Bank Range of Knowledge Correspondence by Grade Band and Domain

Domain	Total Number of PEs	Number of PEs Not Represented	Percentage of PEs Targeted by Items Within Item Bank
Grade 3-5			
PS	17	0	100%
LS	12	0	100%
ESS	13	1	92%
Grade 6-8			
PS	19	1	95%
LS	21	1	95%
ESS	15	0	100%
Grade 9-12			
PS	24	7	71%
LS	24	2	92%
ESS	19	2	89%

Results of the item bank analysis suggest that the item bank has the capacity to assess students across the range of PEs within each domain for the elementary and middle school grade bands, and across the range of the Life Science domain for high school. At least 90% of PEs have one or more corresponding assessment item (stand-alone or cluster) for these grade bands and domains. The high school Earth and Space Science domain was found to include items addressing 89% of the PEs and would be considered to weakly meet the expectations for range, as used in this analysis. The addition of just one item, that addresses one of the two unrepresented PEs would result in the item bank meeting the threshold target of including items for at least 90% of the corresponding PEs. Results suggest that the item bank needs to be supplemented to have the capacity to assess at least 90% of the high school Physical Science PEs as expected by states. To fully meet state expectations, at least five items would need to be added that address five otherwise unrepresented Physical Science PEs. State-specific information within each state-specific report provides evidence as relates to the Range of Knowledge criterion for test events as pertains to the individual as well as to the tested population.

Balance of Representation

Results of the item bank analysis show that items were reasonably distributed among the targeted standards. Between one and eight items were found to correspond to each of the represented PEs. For elementary and middle grades item banks, each PE had an average of approximately three corresponding items/clusters (mean) and most PEs were represented by approximately two items (mode). For the high school item bank, each PE had an average of approximately two corresponding items/clusters (mean) and most PEs were represented by one or two items (mode). No PE was overemphasized in the item bank. Overemphasis typically applies only in circumstances when one particular standard is represented in excess of all others. State-specific information within each state-specific report will provide evidence as relates to the Balance of Representation criterion for test events.

Relationship of Scoring Assertions to Items

Most item clusters had eight corresponding scoring assertions with an average of 10 and up to 20 different assertions, maximum. Most stand-alone items had two corresponding scoring assertions, with an average of two and up to seven different assertions, maximum. Panelists were instructed to carefully read through each individual Scoring Assertion and then consider whether or not the assertions, in aggregate, adequately reflected reasonable inferences about student knowledge, skills, and abilities based on their work on the assessment item or item cluster. Panelists could find that one or more of the assertions slightly misstated, overstated, or understated the inferences that could be made but to code this criterion affirmatively (YES), panelists needed to agree that a

large majority (~75%) of the scoring assertions described a direct inference that could be made from the student's correct responses. Items and item clusters were considered to adequately meet this criterion if a majority (three out of five or four out of six panelists) of a panel coded affirmatively as represented in their independent coding. Results of the item bank analysis show that, with a few exceptions, panelists largely agreed that the Scoring Assertions reasonably described the inferences that could be made based on successful student interactions with the assessment. For just two (out of 127-177) items per grade band, independent coding did not yield a panel majority agreement with the Scoring Assertions. Although panelists thought the alignment considerations were reasonable, they included a variety of notes about the Scoring Assertions with editorial feedback and other comments that could be used to inform continuous improvement.

Relationship of Scoring Assertions to PEs

Panelists also considered the relationship of the Scoring Assertions, in aggregate, to the target PE for an item or item cluster. This evaluative point was intended as a cross-check to “close the loop” on the measurement chain of reasoning for each item or item cluster. If an item (or cluster) adequately targets a particular PE, and the Scoring Assertions appropriately reflect inferences about a student's successful work on the item (or cluster), then it would be expected that the Scoring Assertions circle back to the PE, and adequately reflect at least two (for stand-alone items) or all three (for item clusters) of the three-dimensional expectations therein. Results of the item bank analysis show that, with a few exceptions, panelists largely agreed that the Scoring Assertions reasonably reflected the expectations of the corresponding PE. For three items per high school and middle school grade band, a panel majority did not agree with the Scoring Assertions per independent coding. For the elementary school grade band, a panel majority did not agree with the Scoring Assertions for 11 items, and fewer items were coded affirmatively for this criterion by all panelists in the elementary school grade band compared to the other grade bands.

This study included a high-level content analysis of these assertions to evaluate alignment considerations. CAI provides additional information about the development and use of scoring assertions in Volume 2 of the Foundational Report.

Items Flagged for Revisions or Removal

Panelists were able to map nearly all of the items and item clusters for each grade-band item bank to a specific Performance Expectation. Across all grade bands, only two item clusters, both within the elementary grade band item pool, were identified as not meeting one or more of the alignment expectations. For one of these, the item cluster was considered not to address the internally coded PE, but only because of an error in the internal metadata. Although panelists did

not agree with the internally coded PE but thought the item was a good fit for another PE, and continued analyzing the item with consideration of the presumed corrected PE. This issue could be resolved by reviewing and correcting the internal metadata. For the other item cluster, panelists noted multiple concerns including a mismatch between the Scoring Assertions and the actual student interactions. Across all grade bands, 18 stand-alone items were identified that did not meet one or more of the alignment expectations. Of these, the majority of items (11) were in the elementary grades item bank, six were in the middle school item bank, and only one item was in the high school item pool. Seven of the elementary stand-alone items and one of the high school stand-alone items were considered to inadequately address a PE. Other issues identified included items that offered Category 1 interactions only (which overlaps with unmet Use of Phenomenon), and items with disconnects between the Scoring Assertions and PE or student interactions. All items that did not meet one or more of the state expectations are summarized in **Table 11** and itemized in **Appendix C**. For both middle school and high school the number of items that did not meet one or more of the alignment expectations constituted less than 5% of the overall items in the grade-band item bank. For the elementary grades, approximately 9% of the items in the item bank did not meet one or more of the alignment expectations. While CAI and states may wish to review and revise or remove these items, they do not exceed the cutoffs for acceptable alignment established for this study. In general, states agreed that when an expectation was expected for **all** items, a 90% cutoff would be used to allow some leeway for human error and differences in professional opinion. When aggregating items that did not meet one or more expectation, the proportion of elementary grades items still falls under this threshold.

Table 11. Shared Science Assessment Items with One or More Unmet Alignment Expectations by Grade Band

Grades 3-5 (136 items total in grade-band item bank)		
Issue	# of items (item type)	Suggested Resolution
Inadequate match for PE	7 (stand-alone)	Remove or revise
Category of Engagement Level 1	3 (stand-alone)	
Internally miscoded to wrong domain	1 (item cluster)	Correct internal metadata
Scoring Assertions too far from PE	2 (stand-alone)	Review and edit
Grades 6-8 (177 items total in grade-band item bank)		
Issue	# of items (item type)	Suggested Resolution
Category of Engagement Level 1	4 (stand-alone)	Remove or revise
Scoring Assertions too far from PE	2 (stand-alone)	Remove or revise
Grades 9-12 (126 items total in grade-band item bank)		
Issue	# of items (item type)	Suggested Resolution
Inadequate match for PE	1 (stand-alone)	Remove or revise

Source of Challenge Issues and Panelists' Comments

Panelists were instructed to document any Source of Challenge issue and to provide any other comments they had about an item. A Source of Challenge is a technical issue with an item that can result in a student answering the item correctly or incorrectly for the wrong reason. All of the eight elementary item batches, three of the 12 middle school batches, and five out of eight high school batches contained multiple items for which panelists logged Source of Challenge issues. The main themes from these comments related to items containing expectations beyond grade-level (e.g. requiring use of mathematics from above grade level) or outside of a PE's assessment boundaries, unclear directions, graphics, or response modes, and errors in graphics or text that could affect student responses. Panelists included some comments related to Source of Challenge in their notes as well. For some issues and topics, multiple panelists left similar comments. All comments should be reviewed and considered, including those made by an individual panelist, as one person may have noticed something that others did not. Many of the issues identified have straightforward resolutions, including slight adjustments and corrections to errors, after which the items would be expected to be appropriate and viable. Some of these issues require larger-scale reconsideration of the item or item cluster. Panelists also wrote notes about many items. Some notes included actionable suggestions for item improvements. Panelist notes also contain comments and feedback, including many commendations. These notes may be helpful to identify exemplar items that can be used as models for future item development.

After coding each assessment form, panelists were asked to respond to debriefing questions. Responses to these questions provide qualitative and holistic feedback about the item bank samples and the alignment relationship between the standards and the item bank samples. Many of the responses to the debriefing questions echoed the results from the item-level coding, including a variety of editorial concerns. Panelists also expressed concern with the difficulty level of items. While complex items are expected to be difficult to some degree, difficulty is not influenced solely by complexity. Difficulty can be influenced by the clarity of instructions, the amount of effort required, the number of steps in a problem, the reading demands and density of the materials presented, the extent of opportunities for making errors, whether students have had the opportunity to learn included aspects of items, and other factors. Panelist comments can be used to inform revisions that could help limit excess difficulty in items. CAI item statistics can also be used by states to make decisions about the acceptability of items in terms of difficulty. The full text of panelist comments was provided to states and to CAI but redacted for public release.

Reliability Among Panelists

Panelists engaged in limited adjudication of their data after all panelists finished their coding for an assessment. These discussions were used to identify any mistakes in coding and ensure that the data were entered as intended. Panelists were not required to change their coding after discussion unless they found a compelling reason. The agreement statistics shown in **Table 12** were computed after adjudication. The overall intraclass correlation among the panelists' assignment of Category of Engagement levels to items was high for all analyses. An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the panelists. All of the intraclass correlations were 0.84 or higher.

Table 12. Intraclass and Pairwise Comparisons for Assignment of Category of Engagement by Grade Band and Batch of Items

Grades 3-5		
Batch	Intraclass Correlation	Pairwise Comparison
1 (Panel A)	0.91	0.89
1 (Panel B)	0.85	0.88
2	0.94	0.74
3	0.87	0.71
4	0.91	0.73
5	0.93	0.78
6	0.94	0.95
7	0.92	0.78
8	0.82	0.77
Grades 6-8		
Batch	Intraclass Correlation	Pairwise Comparison
1 (Panel A)	0.95	0.77
1 (Panel B)	0.89	0.64
2	0.95	0.87
3	0.90	0.70
4	0.91	0.67
5	0.85	0.69
6	0.87	0.71
7	0.82	0.68
Grades 9-12		
Batch	Intraclass Correlation	Pairwise Comparison
1 (Panel A)	0.84	0.86
1 (Panel B)	0.90	0.91
2	0.91	0.89
3	0.94	0.87
4	0.95	0.92
5	0.96	0.90
6	0.92	0.90
7	0.96	0.88
8 (Panel A)	0.95	0.88
8 (Panel B)	0.87	0.74

Panelists independently recorded a “Yes” or “No” response to each prompt about the relationship of the scoring assertions, in aggregate, with the actual student interactions and with the corresponding PE. Panelists also independently recorded a “Yes” or “No” response to a prompt about whether the item/cluster met the expectations for Use of Phenomenon. Panelists did not conduct adjudication specific to these evaluative prompts but sometimes discussed these codings in the context of overall discussion about an item. Panelist agreement for their codings of Use of Phenomenon is shown in **Table 13**. Panelist agreement for their codings of scoring assertions’ relationship to student interactions and to PEs is shown in **Table 14**. For each table, the first column shows the percentage of items in each grade band item bank for which all panelists coded the same way, either all coding “Yes” or all coding “No” in response to each evaluative prompt. The middle column shows the percentage of items for which all-but-one panelist agreed. The rightmost column shows the percentage of items for which two or three panelists coded differently than the others. Panelists were very consistent in their coding for Use of Phenomenon, with over 90% agreement for all items. Panelists were also very consistent in their coding for the evaluations of the scoring assertions with no more than one panelist coding differently than the others for between 86% to 98% of items reviewed. The greatest disagreement for the evaluation of scoring assertions was for the elementary grade band relationship of scoring assertions to the PE. In this case, the five panelists were split for 19 out of the 131 items, with two panelists disagreeing (either “Yes” or “No”) with the other three panelists.

Table 13. Panelist Agreement for Rating of Use of Phenomenon by Grade Band

Batch	Use of Phenomenon		
	% of items with 100% panelist agreement	% of items with ~80% panelist agreement	% of items with < ~80% panelist agreement
Grades 3-5	92%	6%	2%
Grades 6-8	92%	7%	1%
Grades 9-12	94%	5%	1%

Table 13. Panelist Agreement for Rating of Scoring Assertions to Student Interactions and to PE by Grade Band

Batch	Scoring Assertion to Student Interaction		
	% of items with 100% panelist agreement	% of items with ~80% panelist agreement	% of items with < ~80% panelist agreement
Grades 3-5	77%	18%	4%
Grades 6-8	82%	13%	5%
Grades 9-12	75%	22%	3%
Batch	Scoring Assertion to PE		
	% of items with 100% panelist agreement	% of items with ~80% panelist agreement	% of items with < ~80% panelist agreement
Grades 3-5	60%	26%	14%
Grades 6-8	82%	14%	4%
Grades 9-12	87%	11%	2%

Conclusions

The WebbAlign team of the non-profit Wisconsin Center for Education Products and Services (WCEPS) facilitated a study in July 2019 to evaluate the alignment of a Shared Science Assessment Item Bank with the Next Generation Science Standards' (NGSS) Performance Expectations (PEs). The item bank is shared by multiple states, 10 of which were involved in the 2019 study. The in-person content analyses were conducted in Denver, CO on July 15-19, 2019 with panelists from all of the 10 states.

Research questions, alignment criteria, and acceptable cutoff levels for these criteria as relate to corresponding science standards and assessments were determined through discussion with state officials, and grounded in analyses of test purpose, construct, and blueprint. State officials were asked to confirm the full set of criteria used and review the acceptable levels proposed for each alignment criterion. States were provided the opportunity to make modifications if warranted. The data collection and reporting for the analyses used the finalized alignment criteria and corresponding cutoffs, appropriate for the context of the states' intents and for the particular context of alignment of science assessments with corresponding NGSS PEs.

Study results suggest that the Shared Science Assessment Item Bank has strong capacity to generate aligned test forms, pending state-specific evidence. The only alignment weakness identified could be fully resolved with the addition of just six items to the high school item bank, targeting unrepresented PEs, to meet states' expectations to have the capacity to address at least 90% of the corresponding PEs in each reporting category.

The research questions used to guide the study design and execution are presented on the following pages, along with the corresponding study findings:

Research Question 1:
1. To what extent do the stand-alone items and item clusters satisfy the measurement target claims (PE and scoring assertions) identified in the CAI metadata? • To what extent does an independent expert panel agree that a student's correct response allows for a reasonable inference about the student's proficiency as relates to the three-dimensional expectations within the identified PE? • To what extent does an independent expert panel agree that the explicit inferences about student performance stated in the scoring assertions can reasonably be made based on student responses to a stand-alone item or item cluster? • To what extent does an independent expert panel agree that the explicit inferences about student performance stated in the scoring assertions reflect the states' measurement target claims (PE) identified in the CAI metadata?

Study results show that the vast majority of the stand-alone items and item clusters satisfy the measurement target claims identified in the CAI metadata. Out of the 440 items/clusters included in the analysis and in this common report, only seven items total (3%) were flagged for revision or removal with the primary issue identified as an inadequate match with a PE. Six of these items were elementary stand-alone items and one was a high school stand-alone item. Across grade bands, only four items (1%) were flagged for revision or removal with the primary issue identified as the Scoring Assertions stretching too far from the PE. All other items met the expectations used in this analysis as pertains to Research Question 1.

Research Question 2:
2. What Category of Engagement (cognitive complexity) is required for successful completion of each interaction within a stand-alone item or item cluster and how does this compare with the Category of Engagement assigned to the corresponding PE?

Out of the 440 items/clusters included in the analysis and in this common report, only five items (1%) were flagged for revision or removal with the primary issue identified related to the Category of Engagement. All five items were stand-alone items. These items were found to include only Category 1 interactions. The remaining 99% of items were found to include opportunities for students to engage at Category 2 and Category 3, corresponding proportionately, overall, to the complexity of the grade-band PEs. The vast majority of items considered Category 3 (97%) were item clusters. The item cluster structure of this assessment, along with the multiple types of information provided within the stimulus and the multiple types of student interactions possible, allowed opportunities for a wide variety of tasks that were considered to require Category 3 Cognitive Engagement. Panelists noted that most student interactions within an item cluster were typically Category 2, but that sometimes at least one of the interactions required students to interweave the components of the tasks such that there was at least one Category 3 interaction. Panelists coded an item cluster (defined as the unit of analysis) to the highest Category of Engagement that it included to ensure that coding captured the full scope of the complexity of an item cluster.

Research Question 3:
3. To what extent do the stand-alone items and item clusters satisfy the claim that the assessment is phenomenon-based?

Assessments that draw from the Shared Science Assessment Item Bank are intended to be phenomenon-based, meaning that items/clusters require students to engage multiple dimensions of the standards (“use science”) to make sense of phenomena. While stand-alone items may be two or three dimensional, they were still expected to require students to use multiple dimensions of a PE to make sense of a phenomenon. With the exception of several items flagged for revision or removal, all items were considered to meet states’ expectations for being phenomenon-based.

Research Question 4:
4. To what extent are state-specific assessment programs likely to generate test events that are aligned with corresponding grade-level academic standards, considering depth and breadth (specified in ESSA) as well as other alignment criteria? • Do the test blueprints and other relevant test specifications and documentation reflect appropriate design to support potential alignment of test events with corresponding grade-level academic standards? • Do the available aggregate data for recently administered test events in each state provide evidence that the algorithm and blueprints yielded test forms as expected? • To what degree are actual test events for each state (if available) aligned with corresponding grade-level academic standards for each state?

Research Question 4 is addressed in state-specific reports. To determine the capacity of alignment of the Shared Science Assessment Item Bank to yield test forms aligned with a particular state's context, several additional checks are required. If a state's test blueprint reflects appropriate design to support potential alignment of test events and if a state's aggregate data from all administered test events show that the blueprints and algorithm are yielding test forms as expected, then it is possible—based on the results of the item-level content alignment analysis—to make an argument for the capacity of alignment for all test events resulting from the state's summative science assessment program that use items from the Shared Science Assessment Item Bank. Results from analyses of actual sample test events provide additional state-specific alignment information.

Nine alignment criteria were identified for use in the analysis:

1. **Use of Phenomena:** Each stand-alone item and item cluster is expected to be grounded in a stimulus that meets the test development criteria for a phenomenon. Items/clusters are expected to require students to engage multiple dimensions of the PEs (“use science”) to make sense of those phenomena.
2. **Categorical Concurrence:** Test events are expected to yield sufficient evidence to make inferences about student knowledge, skills, and abilities (KSAs) as relates to each reporting category.
3. **Dimensionality (Structure of Knowledge):** All item clusters are expected to require students to demonstrate integrated engagement with the three dimensions of Science and Engineering Practices (SEPs), Disciplinary Core Ideas (DCIs), and Crosscutting Concepts (CCCs) specified in the targeted PE. All stand-alone items are expected to require students to demonstrate integrated engagement with two or three of the dimensions specified in the targeted PE.
4. **Consistency of Cognitive Engagement:** The assessment is expected to elicit work that is as cognitively demanding as the expectations in the PEs.
5. **Range of Knowledge Correspondence (Individual):** Test events are expected to assess an appropriate breadth of the standards. For individual students, assessed PEs are sampled across topics within each reporting category.
6. **Range of Knowledge Correspondence (Population):** At least 90% of PEs within a grade band have the potential to be assessed across the student population. State-specific claims are consistent with aggregate data from all administered test events in the state in conjunction with results from an independent analysis of vendor metadata.
7. **Balance of Representation:** No PE is targeted more than once on any single test event.
8. **Relationship of Scoring Assertions with Student Interactions:** In aggregate, the scoring assertions for an item/item cluster appropriately represent the inferences about student knowledge, skills, and abilities that can be made based on successful interactions with an item/cluster.
9. **Relationship of Scoring Assertions with PEs:** In aggregate, the scoring assertions for an item/item cluster appropriately represent the three-dimensional expectations of the targeted PE.

Study results suggest that the overall Shared Science Assessment Item Bank has the capacity to fully or weakly meet the criterion of Range of Knowledge Correspondence (Population) and the capacity to fully meet all other alignment criteria used in this study and itemized above. The criteria of Categorical Concurrence, Range of Knowledge Correspondence (Individual), and Balance of Representation is addressed primarily in the state-specific reports, considering state-specific documentation, test events, and data. In general, full or acceptable alignment was found between the NGSS PEs and the assessment items and item clusters for all grade bands although specific items were identified that did not meet one or more expectations, and warrant revisions or removal. Details about these items were provided to states and CAI.

The only alignment weakness identified in the item-level analysis of the Shared Science Assessment Item Bank was that the high school item pool did not meet states' expectations to have the capacity to address at least 90% of the corresponding PEs. This issue could be fully resolved with the addition of at least six items to the high school item bank. At least five of items would need to address unrepresented PEs within the Physical Science domain. At least one item would need to address one of the two unrepresented PEs within the Earth and Space Science domain.

Although panelists found that items and item clusters were meeting state expectations as relates to alignment, they made many editorial suggestions and identified many issues that could broadly be described as editorial. The design of the items and item clusters in this program include elaborate stimuli made up of narratives, tables, graphs, and other visuals, extensive student interactions that include additional text, tables, graphs, and other visuals. The scope of material incorporated into each item and item cluster means abundant opportunities for typos, confusing wording, layout issues, mislabeling, and other editorial issues. Panelists commented on a multitude of corrections and other editorial problems that have implications for student responses and scoring and therefore require or should be considered for revisions. Some of these issues, including errors, typos, and line edits, could be resolved very simply, with minor edits or corrections. Other issues require some reworking of graphics or layout. Panelists also expressed some concerns with aspects of scoring and functionality of interactions, and questioned the underlying logic or accuracy related to some interactions.

CAI may wish to develop more robust style guides or revise existing style guides to ensure that there are comprehensive defined guidelines and conventions for use of terminology, consistency in the use and corresponding intended meaning of visual representations (e.g. arrows), guidelines for formatting and labeling of features including tables, graphs, and diagrams, consistency in the overall structure and contents of directions for items, and other attributes of the

assessment items. While fairly limited style guides have sufficed for more traditional on-demand assessments, the item cluster design incorporates a wide variety of graphics and text, and, consequently, requires much more extensive editorial guidelines and reviews.

In their notes, panelists also documented some inherent conflicts they perceived between certain PEs and their subcomponent dimensions. For example, for the PE 5-ESS1-1, the foundation box for CCC “Scale Proportion and Quantity” specifically relates to size: “Natural objects exist from the very small to the immensely large.” However, the assessment boundary for the PE specifically excludes size and restricts assessment to relative distance. For MS-ESS1-3, the DCI lists the components of the solar system and specifies that they are “held in orbit around the sun by its gravitational pull on them.” The PE, however, does not attend to the concepts of orbit or gravitational pull. Instead, the PE focus is on “scale properties of objects in the solar system.” While orbital radius is listed as an example of a scale property, the overall emphasis is on other physical features of celestial objects. The PE MS-LS1-5 includes a DCI element, which states “Genetic factors as well as local conditions affect the growth of the **adult plant**” (emphasis added). However, the PE refers to “growth of organisms” and the Clarification Statement gives examples that include animals as well as plants. Panelists noted that an item addressing a non-plant organism would therefore seem to be appropriate, based on the language of the PE, despite the DCI element specifying plants only. In general, panelists allowed flexibility for the evaluation of item dimensionality for instances where they perceived internal inconsistencies between specific elements of the dimensions and the corresponding PE.

Overall, a content alignment analysis of all stand-alone items and item clusters from the Shared Science Assessment Item Bank (available at the time of the study) shows that the items/clusters were appropriate as relates to intended claims and inferences. Panelists found that items and item clusters were meeting state expectations that assessment tasks required integrated engagement with at least two dimensions (for stand-alone items) or all three dimensions (for item clusters) of SEPs, DCIs, and CCCs specified in the targeted PE in order to make sense of a phenomenon. Study results suggest that items required student cognitive engagement consistent with the expectations of the corresponding standards. With the exception of the high school Physical Science item pool, Range was considered met or weakly met, with one or more item(s) or item cluster(s) that represented all or all but one or two PEs within each grade band and domain. Unmet Range for the high school item bank could be fully resolved with the addition of at least six items, each targeting an unrepresented PE (five of which are Physical Science PEs). Item were spread across the domains of Physical, Life, and Earth and Space Science, with no PE(s) overemphasized in the item bank. Overall, panelists found that the large majority of scoring

assertions reasonably reflected inferences that could be made based on student interactions and corresponded to the expectations within the targeted PE. While alignment results were favorable, the extent of errors and editorial issues found in text and graphics is typically not observed in an operational assessment and merits attention. After noting some of these editorial issues, one panelist summarized, “Overall, the items seem strong and do a commendable job of assessing proficiency as it relates to the 3D standards.”

State-specific reports, considering state-specific documentation, test events, and data, provide additional evidence that can be used to judge the capacity for alignment of all test events generated by the Shared Science Assessment Item Bank with corresponding state standards.

References

- Achieve, Inc.; Stanford/SCALE (2019) Reconceptualizing Alignment for NGSS Assessments. NARST Annual Meeting. Symposium paper available (01/2020) at https://snappgse.stanford.edu/sites/g/files/sbiybj10126/f/narst_symposium_2019_3.31.19.pdf
- Badrinarayan, A., Christopherson, S. and Harris, C. (2017) The Alignment Challenge. Reidy Interactive Lecture Series: Assessing Student Learning of the Next Generation Science Standards. The National Center for the Improvement of Educational Assessment
- Badrinarayan, A, Christopherson, S., Gong, B., McCrae, A. (2018) Developing a Common Language to Understand Content Complexity for Alignment Studies of the NGSS. National Conference on Student Assessment
- Badrinarayan, A., Christopherson, S., Davis-Becker, S., Everson, H., and Forte, E. (2019). Representing Cognitive Complexity in Test Design and Evaluation. ATP Innovations in Testing
- National Research Council. 2012. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press. <https://doi.org/10.17226/13165>.
- National Research Council. (2005). How Students Learn: History, Mathematics, and Science in the Classroom. Committee on How People Learn, A Targeted Report for Teachers, M.S. Donovan and J.D. Bransford, Editors. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- NGSS Lead States. 2013. Next Generation Science Standards: For States, By States. Washington, DC: The National Academies Press.
- NGSS Lead States. 2013. NGSS Release: How to Read the Next Generation Science Standards (NGSS). Washington, DC: The National Academies Press.
- Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25(1), 47-55.
- Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literary standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III*. Mahwah, NJ: Lawrence Erlbaum.

- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Council of Chief State School Officers and National Institute for Mathematics Education Research Monograph No. 6. Madison: University of Wisconsin, Wisconsin Center for Education Research.
- Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison, WI: University of Wisconsin.
- Webb, N. L. (1999). Alignment of science and mathematics standards and assessments in four states. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 18. Madison, WI: University of Wisconsin.
- Webb, N. L. (2003). Alignment study in language arts, mathematics, science, and social studies of state standards and assessments in four states. Washington, D. C.: Council of Chief State School officers.
- Wise, S. L., Kingsbury, G. G., and Webb, N. L. (2015). Evaluating content alignment in computerized adaptive testing. *Educational Measurement: Issues and Practices*, Winter, 34, 4, pp. 41-48.
- CT Technical Report Connecticut State Department of Education / American Institutes for Research (2019). Connecticut Alternate Science Assessment Standard Setting Technical Report.

**Cambium Statewide Science Assessment Item Bank
for States Sharing Science Content**

State-Specific Report: West Virginia

**Alignment Analysis of the West Virginia General Summative
Assessment (WVGSA) for Science Grades 5 and 8 with
Corresponding West Virginia Next Generation Content Standards and
Objectives for Science**

Sara C. Christopherson and Norman L. Webb
June 18, 2020

Wisconsin Center for Educational Products and Services
Matt Messinger, Executive Director
University of Wisconsin Research Park
510 Charmany Drive, Suite 269
Madison, WI 53719

This report applies to the state of West Virginia and should be considered in the context of the results of the independent third-party alignment analysis of the Shared Science Assessment Item Bank, detailed in a common report for the participating states of Connecticut, Hawai'i, Idaho, New Hampshire, Oregon, Rhode Island, Utah, Vermont, West Virginia, and Wyoming. The overall study was coordinated and funded by Cambium Assessments (CAI; formerly American Institutes for Research – AIR). Toni Deoudes was the main contact for communication with CAI. A Working Group comprised of state leadership also contributed to study planning and coordination. Timothy Butcher was the main contact for communication with West Virginia.

Table of Contents

Table of Contents

Executive Summary	1
Introduction and Methodology.....	3
Alignment Criteria Used for This Analysis.....	5
Cutoffs for Each Criterion for Individual Test Events	7
Cutoffs for Overall Alignment of Test Events with PEs	7
Standards, Reporting Categories, and Blueprints.....	8
West Virginia Science Item Bank for Grades 5 and 8.....	10
Alignment Findings: Sample West Virginia Test Forms.....	11
Alignment Findings: Overall West Virginia General Summative Assessment for Science Program Capacity to Generate Aligned Assessments	17
Conclusion	24
References.....	27

Executive Summary

This report describes the results of a content alignment analysis of the West Virginia General Summative Assessment (WVGSA) for science for grades 5 and 8 with corresponding West Virginia Next Generation Content Standards and Objectives for Science (equivalent to the Next Generation Science Standards) as pertains to fulfilling requirements as stated in Federal statute. The West Virginia assessment used a particular state-vetted subset of items that were part of a Shared Science Assessment Item Bank. The item bank is managed by Cambium Assessment (CAI; formerly known as American Institutes for Research - AIR) and is shared by multiple states.

An item- and item-cluster-level content alignment analysis was completed in July 2019 to evaluate the alignment of the overall Shared Science Assessment Item Bank with the Next Generation Science Standards' (NGSS) Performance Expectations (PEs). Participants from 10 of the states that use the shared item bank were involved in the 2019 study. Participating states agreed to share items through a Memorandum of Understanding (MOU) that details a commitment to shared content, leadership, ideas, and methods. A common report describes the overall results of the alignment analysis of the item bank, applicable to all states.

Results of the item-level analysis showed, in general, full alignment between the NGSS PEs and the assessment items and item clusters for elementary (grade 3-5) and middle school (grade 6-8) grade bands. However, specific items were identified that did not meet one or more expectations, and warrant revisions or removal (details provided in the common report).

Although study results suggest that the Shared Science Assessment Item Bank has strong capacity, overall, to generate aligned test forms, state-specific evidence is also necessary for a judgement of the capacity to generate aligned test forms for a particular state context. The overall results of the item-level content analyses were applied to West Virginia's assessment framework and statistical documentation, in the context of the state-specific item subset, to make some inference about the capacity for alignment of all test events generated by the WVGSA for science for grades 5 and 8 with corresponding West Virginia Next Generation Content Standards and Objectives for Science. Three West Virginia sample test events for each of grades 5 and 8 provided test-event-level evidence for the state. This report includes West Virginia-specific details and applies to the state of West Virginia only.

The results described in this West Virginia-specific report include test-event-level findings followed by findings related to the overall capacity for alignment of the WVGSA for science in terms of all nine criteria agreed upon by participating states to be used to evaluate alignment of the assessments with corresponding NGSS PEs, which are equivalent to the West Virginia Next Generation Content Standards and Objectives for Science. The overall findings reported are grounded in a synthesis of the item-level, test-event-level, and item-bank-level findings.

The West Virginia General Summative Assessment (WVGSA) for science grades 5 and 8 was found to be acceptably aligned with the corresponding state standards for science, meaning that it had the overall capacity to generate test forms that were all fully or acceptably aligned with the West Virginia Next Generation Content Standards and Objectives for Science (equivalent to the NGSS). This finding includes consideration of the results of the item-level analyses of the overall item bank, the West Virginia item bank, sample West Virginia test events, West Virginia blueprints and aggregate data from all administered West Virginia test events. The evidence to support this finding includes:

- The West Virginia test blueprints identified the state's intended sampling across reporting categories (as relates to Categorical Concurrence, Range of Knowledge (breadth) for individual test events, and Balance of Representation (emphasis)).
- Aggregate data from all administered test events within the state of West Virginia showed that the blueprints and item selection algorithm yielded test forms as expected.
- Overall, the items within the West Virginia item bank met the state's expectations (as relates to Use of Phenomena, Dimensionality, Consistency of Cognitive Engagement, and relationships with scoring assertions).
- The West Virginia item bank was found to fully meet the state's expectations for Range of Knowledge across the tested student population for all grades and domains.
- An analysis of three sample test events from each of grades 5 and 8 found that all test events were fully or acceptably aligned with corresponding standards, based on the criteria agreed upon by states and used in this analysis.

Introduction and Methodology for Content Alignment Analyses

An overall description of the in-person content alignment analyses conducted in July 2019 is provided in the common report and is not repeated here. The common report addresses the first three research questions that guided the study and that apply to the item-level analyses. This report primarily addresses the fourth, and ultimate, research question that guided the overall study, shown in boxed text below. This ultimate research question involves consideration of state-specific item banks and documentation in the context of the greater item-bank-level findings as well as consideration of sample test events.

Research Question 4:
1. To what extent are state-specific assessment programs likely to generate test events that are aligned with corresponding grade-level academic standards, considering depth and breadth (specified in ESSA) as well as other alignment criteria? • Do the test blueprints and other relevant test specifications and documentation reflect appropriate design to support potential alignment of test events with corresponding grade-level academic standards? • Do the available aggregate data for recently administered test events in each state provide evidence that the algorithm and blueprints yielded test forms as expected? • To what degree are actual test events for each state (if available) aligned with corresponding grade-level academic standards for each state?

The overall alignment study was structured to allow for the potential to build a logic argument for the capacity for alignment of all test events generated by the Shared Science Assessment Item Bank with corresponding state standards, as appropriate, based on results. As such, the study was designed to generate multiple lines of evidence that could be used to support a claim that the item bank had the capacity to yield aligned test forms for each state. This evidence comes from study findings at the item-level, test-event-level, and item-bank-level. Several conditions must be met for such a claim to be supported:

- A state's test blueprint must have reflected appropriate design to support potential alignment of test events, according to the state's intended claims about the assessment (as relates to Categorical Concurrence, Range of Knowledge (breadth) for individual test events, and Balance of Representation (emphasis)).
- Aggregate data from all administered test events within the state must show that the blueprints and item selection algorithm yielded test forms as expected.

- The items within the state's item bank must have met the state's expectations (as relates to Use of Phenomena, Dimensionality, Consistency of Cognitive Engagement, and relationships with scoring assertions)
- A content alignment analysis of sample test events for each grade must have found full or acceptable alignment of test events with corresponding standards, based on the criteria agreed upon by states and used in this analysis.

If these conditions were met, then it was possible to make an argument for the capacity of alignment for all test events resulting from a state's summative science assessment program that used items from the Shared Science Assessment Item Bank. Results from the item-level analyses of the overall item bank were mapped onto the item sets from three West Virginia sample test events for each of grades 5 and 8 to provide state-specific test-event-level alignment information. Alignment results at the test-event level are given within the Findings section of this document followed by overall alignment findings for the WVGSA for science for grades 5 and 8.

The results reported here pertain only to the issue of alignment between the West Virginia standards and assessments. Note that an alignment analysis of this nature does not serve as external verification of the general quality of the standards or assessments, but rather, the focus is on the degree of alignment. Throughout this document, the term "item" may refer to both stand-alone items and to item clusters.

Alignment Criteria Used for this Analysis

After input from and discussion with representatives from all states, the nine alignment criteria detailed in **Table 1** were agreed upon to be used to evaluate and report on the degree of alignment of standards with state assessments, including the WV GSA for science for grades 5 and 8 with corresponding state science standards. Details on these criteria are provided in the shared report and are not repeated here.

Table 1. Consensus Alignment Criteria for Shared Science Assessment Item Bank and State Assessment Programs, 2019

Criterion	Intended Claim/Inference	Acceptable Cutoff
1. Use of Phenomena	Items/clusters require students to engage multiple dimensions of the standards (“use science”) to make sense of phenomena. Each assessment item/item cluster is grounded in a stimulus that meets the test development criteria for a phenomenon.	At least 90% of items/clusters are considered phenomenon-based by a majority of panelists (i.e. at least 3 out of 5 or 4 out of 6 panelists).
2. Dimensionality/ Structure of Knowledge Comparability	Item clusters require students to demonstrate integrated engagement with the three dimensions of SEPs, DCIs, and CCCs specified in the targeted PE. Stand-alone items require students to demonstrate integrated engagement with two or three of the dimensions specified in the targeted PE.	At least 90% of clusters are considered three-dimensional by a majority of panelists; at least 90% of stand-alone items are considered multi-dimensional by a majority of panelists.
3. Categorical Concurrence*	Test events have the potential to yield sufficient evidence to make inferences about student knowledge, skills, and abilities (KSAs) as relates to each reporting category.	A test form will include at least six (6) opportunities to respond to items or clusters that target the standards within each reporting category.
4. Consistency of Cognitive Engagement*	The assessment elicits work that is as cognitively demanding as the expectations in the standards.	While some individual student interactions may be Category 1, no items/clusters should include only Category 1 interactions. Proportions of items/clusters with Category 2 and 3 opportunities reflect grade band PEs. Some aspects of Category 4 PEs will be assessed but the full scope of Category 4 PEs is expected to be assessed in the classroom.

Table 1 Cont'd. Consensus Alignment Criteria for Shared Science Assessment Item Bank, 2019

Criterion	Intended Claim/Inference	Acceptable Cutoff
5. Range of Knowledge Correspondence (Population)*	State-specific claims will be considered against aggregate data from all administered test events in the state in conjunction with external confirmatory analysis of vendor metadata. Confidence in vendor metadata depends on results of external confirmatory analysis.	At least 90% of PEs have the potential to be assessed across the student population.
6. Range of Knowledge Correspondence (Individual)*	Test events assess an appropriate breadth of the standards, as defined by the state assessment blueprint. Assessed standards are sampled across topics within each reporting category for individual students.	Test forms analyzed meet blueprint specifications for Range. Blueprints are expected to specify sampling across topics (or other sublevels for each reporting category).
7. Balance of Representation*	No PE is targeted more than once on any test event.	A PE should not be targeted more than once on a test event; each stand-alone item and item cluster should target a different PE.
8. Relationship of Scoring Assertions with Student Interactions	In aggregate, the scoring assertions for an item/item cluster appropriately represent the inferences about student knowledge, skills, and abilities that can be made based on successful interactions with an item/cluster.	For at least 90% of all items/clusters, a majority of panelists consider a large majority of the scoring assertions (at least ~75%) to appropriately represent the inferences about student KSAs that can be made based on successful interactions with an item/cluster.
9. Relationship of Scoring Assertions with PEs	In aggregate, the scoring assertions for an item/item cluster appropriately represent the three-dimensional expectations of the targeted PE. (At least two of the three dimensions for stand-alone items.)	For at least 90% of all items/clusters, a majority of panelists consider a large majority of the scoring assertions (at least ~75%) to appropriately represent the expectations within the corresponding PE.

*West Virginia-specific results for these criteria are included in this report.

Cutoffs for Each Criterion for Individual Test Events:

For individual test events, acceptable alignment for Categorical Concurrence, Range of Knowledge (Individual), and Balance of Representation is defined by the blueprint. Categorical Concurrence and Range of Knowledge (Individual) are considered met if the test form is consistent with the cutoffs used in this analysis (**Table 1**), weakly met if the test form falls short by no more than one item cluster or two stand-alone items per criterion and domain, and unmet if the test form falls short by more than one item or two item clusters per criterion and domain. The Balance of Representation criterion for test events is binary and is either met or unmet.

For individual test events, a reporting category was considered to have met the criterion of Consistency of Cognitive Engagement if the domain had no stand-alone items or item clusters with only Category 1 interactions and reflected the distribution of complexity as expressed in the PEs, meaning the domain included at least one item with one or more interactions that required Category 3 cognitive engagement. This expectation was considered weakly met if results fell short by no more than one item per domain. Weakly met indicates that the criterion was nearly met, within a margin that could simply be due to error or reasonable variation in reviewer coding. The criterion was considered unmet if results fell short by two or more items per domain.

Cutoffs for Overall Alignment of Test Events with PEs:

Typically, a summative assessment test form has been considered fully aligned with corresponding standards if no changes were needed and acceptably aligned if it needed between one and five items revised or replaced. This widely accepted decision rule was grounded in the context of a typical multiple-choice test form of around 50 items that were generally equally weighted. Five items therefore constituted approximately 10% of the test form. If between six and 10 items (more than 10% and up to 20% of items) needed revision or replacement, the test form was considered to need slight adjustments. If a test form needed over 10 items (greater than 20% of items) revised or replaced, it was considered to need major adjustment. These decision rules do not apply in the context of the WVGSA for science, which includes multi-part items and item clusters that vary in the number of associated scoring assertions. Stand-alone items have an average of two scoring assertions and up to seven scoring assertions, maximum. Item clusters have an average of 10 scoring assertions and up to 20 scoring assertions, maximum.

Because items vary in their contribution to a student's score, the approximate percentage of scoring assertions affected by unmet alignment expectations was used to categorize the degree of alignment for a test event. The same typical

decision rules were applied, as described on the previous page, but in the context of the scoring assertions. Therefore, a test form was considered fully aligned if no changes were needed and acceptably aligned if it needed revisions or replacements corresponding to up to 10% of the overall scoring assertions for the test form. A test form was considered to need slight adjustments if it needed revisions or replacements corresponding to between 10% and 20% of the overall scoring assertions for the test form, and to need major adjustments if it needed revisions or replacements corresponding to over 20% of the overall scoring assertions for the test form.

To determine the overall percent of affected scoring assertions, the exact number of scoring assertions was used for any specific item(s) that required revision or replacement. For test forms that needed the addition of one or more items that offered Category 3 cognitive engagement, a per-item estimate was used of 4% (grade 5) and 3% (grade 8) of the total scoring assertions for the test form. These estimates are based on an average item cluster, comprised of three parts and associated with 10 scoring assertions, and an average test form, associated with 82 scoring assertions (West Virginia, grade 5) or 96 scoring assertions (West Virginia, grade 8). When an item cluster was rated a Category 3, the opportunity for Category 3 cognitive engagement generally corresponded to one part of a multi-part item. Therefore, inclusion of a Category 3 opportunity can be considered equivalent to at least around 3 scoring assertions, approximately 4% (grade 5) or 3% (grade 8) of the total scoring assertions for a test. These decision rules allow for an overall categorization of the degree of alignment that take into account the varying contribution of items to a student's score. For example, a test form would be found to need slight improvements if it needed replacement or revision of three items, each of which was associated with multiple scoring assertions such that the total number of assertions was >10% of the overall set of assertions for the test form. Another test form, however, would be found to be acceptably aligned if it needed replacement or revision of three items, each of which were associated with a single scoring assertion, as in that case the overall proportion of affected scoring assertions would only be approximately 3-4%.

Standards, Reporting Categories, and Blueprints:

The West Virginia Next Generation Content Standards and Objectives for Science were adopted by the West Virginia Board of Education in 2015. No adjustments were made from the language of the Next Generation Science Standards for the context of West Virginia, except for the replacement of the word “rise” with “change” in the following expectation: S.6.ESS.6 (MS-ESS3-5) *Ask questions to clarify evidence of the factors that have caused the change in global temperatures over time.*

West Virginia science assessment blueprints for grades 5 and 8 use the domains (reporting categories) of Life Sciences, Physical Sciences, and Earth and Space Sciences and study results are reported according to these categories. Each domain is further divided into sub-domains according to the NGSS DCI arrangement of the PEs.

Blueprints specify the length of the test and the minimum and maximum number of stand-alone items and item clusters that can be included on a test event by DCI organization of the PEs per domain. Blueprints specify the number of item clusters and stand-alone items administered per domain and the maximum total items that can be sampled from the same DCI (no more than one item cluster and two stand-alone items). Blueprints also specify that each PE is represented on a test event by no more than one item cluster or stand-alone item and that items are selected from across DCIs. Blueprint specifications take multiple factors into consideration, including item distribution within reporting categories, exposure effects, and the depth of the item pool for each reporting category and DCIs within.

The final blueprint design, therefore, reflects the West Virginia Department of Education's (WVDE) expectations for the assessment of individual students, based on thoughtful deliberation and examination of multiple factors. As such, the grade-level blueprints are the referents, as defined by WVDE, for a consideration of alignment as pertains to the criteria of Categorical Concurrence, Range of Knowledge (breadth) for an individual student, and Balance of Representation.

Blueprints do not specify a distribution of the Category of Engagement (i.e. type of cognitive engagement, also commonly called “cognitive complexity” or “depth”) of items. However, item-level findings supported the expectation that all items within the item bank (with just a few exceptions) require Category 2 or 3 engagement and appropriately reflect the overall distribution of cognitive engagement expressed within in the corresponding PEs.

Blueprints do not specify dimensionality. However, item-level findings supported the expectation that item clusters within the item bank require students to demonstrate integrated engagement with all three dimensions (SEP, DCI, and CCC). Similarly, item-level findings supported the expectation that stand-alone items require students to demonstrate integrated engagement with at least two of the three dimensions (typically, DCI + SEP).

Test events were administered online and used linear-on-the-fly (LOFT) item delivery in which the item selection algorithm chooses items based solely on content value toward blueprint fulfillment. In other words, each item is selected based on its contribution to meeting the blueprint specifications, given the items that have already been administered.

West Virginia Science Item Bank for Grades 5 and 8:

The West Virginia science item bank includes items owned by West Virginia, items shared from the AIRCore item bank, and items shared by other states that all signed a Memorandum of Understanding (MOU). The MOU details a commitment to share content, leadership, ideas, and methods, including following the same item specifications, test development processes, and review processes. All items from the West Virginia item bank available at the time of the July 2019 study were included in the analysis and shown in **Table 2** by grade and domain. To identify the subset of items used in the West Virginia item bank, the item Master Identification Numbers from Appendix C of the WVGS 2018-2019 Technical Report Volume 1 were compared against the Master Identification Numbers from the items analyzed in the item-level alignment study. Any items from the overall item pool that did not pertain to West Virginia were removed, yielding the subset of items that corresponds to the state of West Virginia. In the time since July 2019, some items were added and some were removed from the item pools. The results reported here are based on all items included in the analysis that are identified as West Virginia items as represented in Appendix C of the assessment technical report.

Table 2. Number of Shared Science Assessment Items Used by West Virginia and Included in this Report by Domain and Grade

Number of Items by Grade and Reporting Category		
	Grades 5	Grades 8
Physical Science	41	44
Life Science	41	60
Earth and Space Science	35	44
TOTAL	117	148

Alignment Findings: Sample West Virginia Test Forms

Results from the item-level analyses were mapped onto the item sets from three actual West Virginia sample test events for each of grades 5 and 8 to provide additional state-specific alignment information. Based on the results of the item-level analysis, stand-alone items are essentially fungible, as all items (with specific exceptions) met the expectations for Use of Phenomenon, Dimensionality (Structure of Knowledge) as related to the PE the item addressed, and for the relationships of scoring assertions to student interactions and to the PE. Stand-alone items were also nearly all found to require Category 2 cognitive engagement. Item clusters can be expected to meet all of the same criteria with the exception of some variation in the Category of Cognitive Engagement required. Some item clusters were considered Category 2 while others were considered to include at least one opportunity for cognitive engagement at Category 3. Therefore, the sample test forms can be considered largely representative of West Virginia test forms, overall.

The West Virginia science assessment for each of grades 5 and 8 consists of 18 items: six item clusters and 12 stand-alone items. All interactions are machine scorable. Information about item types and structure is found in the common report and is not repeated here.

Overall test-event-level alignment results are summarized in **Table 3**. Based on the cutoffs for the alignment criteria agreed upon by states and used in this study, all test forms were found to be fully or acceptably aligned with corresponding grade-band West Virginia Next Generation Content Standards and Objectives for Science. The approximate numbers of replaced or revised items necessary for full alignment are provided for each test form. Because items vary in their contribution to a student's score, the approximate percentage of affected scoring assertions was used to categorize the degree of alignment.

Table 3. Overall Alignment Findings for West Virginia Science Grades 5 and 8 Sample Test Forms with Corresponding Grade Band NGSS

Test Form	Findings	Approx. Number of Items that Need Revision/ Replacement for Full Alignment	Approx. % of Total Assertions that Need Revision/ Replacement for Full Alignment
WV Grade 5 Form 1	Fully Aligned	--	--
WV Grade 5 Form 2	Acceptably Aligned	3 items	7%
WV Grade 5 Form 3	Acceptably Aligned	3 items	10%
WV Grade 8 Form 1	Fully Aligned	--	--
WV Grade 8 Form 2	Acceptably Aligned	3 items	10%
WV Grade 8 Form 3	Acceptably Aligned	1 item	4%

The criteria of Use of Phenomenon, Dimensionality, Consistency of Cognitive Engagement, and the Relationship of Scoring Assertions with Student Interactions and with PEs were met, overall, at the item- and item-bank level. Specific items that did not meet these criteria and that were included on the sample test events were taken into consideration in the overall alignment results reported here. In other words, if a test event included one of the items flagged for removal in the overall item bank analysis, that item was considered to need revision or replacement for full alignment as reported in the results. For the test-event level, the criteria of Categorical Concurrence, Consistency of Cognitive Engagement, Range of Knowledge for the individual test event, and Balance of Representation also require attention. Range of Knowledge for the overall tested population is considered at the level of the item bank.

For Categorical Concurrence, Range of Knowledge (for the individual test event), and Balance of Representation, acceptable alignment is defined by the West Virginia blueprints. Because aggregate data from all administered test events in West Virginia show that essentially all test events (only very few exceptions; <0.0001% of tests) met blueprint specifications, all test events therefore fully met these criteria with the exception of tests events that contained one or more item(s) that were flagged with issues in the item-level analysis.

The criterion of Consistency of Cognitive Engagement expects the assessment to elicit work that is as cognitively demanding as the expectations in the standards. Based on state expectations, this means that while some individual student interactions may be Category 1, no items/clusters should include only Category 1 interactions. Proportions of items/clusters with Category 2 and 3 opportunities within an item bank are expected to reflect the proportion of Category 2 and 3 expectations within the grade band PEs. (Some aspects of Category 4 PEs may be assessed but the full scope of Category 4 PEs is

expected to be assessed in the classroom.) Because the item bank was found to contain an appropriate distribution of items with Category 2 and 3 cognitive engagement, corresponding to the expectations within the West Virginia Next Generation Content Standards and Objectives for Science, Consistency of Cognitive Engagement was found to be met at the overall level of the assessment program.

For individual test events, a reporting category was considered to have met the criterion of Consistency of Cognitive Engagement if the domain had no stand-alone items or item clusters with only Category 1 interactions and reflected the distribution of complexity as expressed in the PEs, meaning the domain included at least one item with one or more interactions that required Category 3 cognitive engagement. This expectation was considered weakly met if results fell short by no more than one item per domain and unmet if results fell short by two or more items per domain.

Because each test form addresses a different set of PEs, and because the West Virginia assessment blueprints do not select for cognitive engagement (i.e. “cognitive complexity” or “depth”) the distribution of items at Category 2 and Category 3 cognitive engagement was expected to vary to some extent between and among test events. Therefore, this variation is not considered an alignment issue based on state expectations. The findings related to the distribution of items by Category of Engagement at the test event level are reported here for WVDE informational purposes.

Based on the item-level analysis of sample test events, all reporting categories (domains) for all grades fully met the criteria for Categorical Concurrence, Range of Knowledge for the individual test event, and Balance of Representation. For all sample test events, all reporting categories for all grades fully met or weakly met the criterion of Consistency of Cognitive Engagement with the exception of the Physical Science domain on one of the grade 8 test forms. The distribution of items by Category of Cognitive Engagement is shown in **Tables 4** and **5**. Any alignment weaknesses identified for test forms are described along with key qualitative feedback from panelists.

Grade 5 Sample Test Events

The three West Virginia grade 5 sample test events analyzed were found to be fully or acceptably aligned based on the criteria used in this analysis. To fully meet state expectations, Forms 2 and 3 were found to need three items revised to ensure no items with only Category 1 interactions and at least one item per domain that included one or more interactions requiring Category 3 cognitive engagement. On these forms, three items (1474 on both forms, 1564 on Form 2), were stand-alone items associated with a single scoring assertion that were considered to include only Category 1 cognitive engagement.

Panelists expressed some concerns about the science content as related to the PEs for two stand-alone items on the test forms. Although panelists agreed the items were phenomenon-based and multidimensional, addressed core aspects of a PE, and required Category 2 or 3 cognitive engagement, panelists suggested revisions for better correspondence to PEs. Item 1543 (Form 2; a stand-alone item associated with two scoring assertions) addressed the PE 5-LS1-1: *Support an argument that plants get the materials they need for growth chiefly from air and water*. While the item directly addressed water, panelists commented that the item did not explicitly attend to air. Because most plant biomass comes from air (carbon dioxide) panelists were concerned that the focus on water over air had the potential to perpetuate common misconceptions related to plant growth. One item (1581; a stand-alone item associated with a single scoring assertion) was included on all three forms and addressed 3-ESS3-1: *Make a claim about the merit of a design solution that reduces the impacts of a weather-related hazard*. Panelists commented that although the circumstance could apply to a weather-related hazard, there was no weather-related hazard specified in the phenomenon provided and they suggested some revisions to strengthen the connection with the language of the PE. Panelists left comments and suggested revisions for item 1623, a stand-alone Earth and Space Science item associated with a single assertion. Panelists included qualitative feedback and suggestions for many other items across forms that merit consideration.

Overall, however, for Form 1, all domains fully met all alignment criteria. For Forms 2 and 3, all domains weakly met Consistency of Cognitive Engagement, and fully met all other criteria.

Table 4. Distribution of Items by Category of Cognitive Engagement, West Virginia Grade 5 Forms 1-3

Grade 5 Form 1	Cognitive Engagement by Domain		
Reporting Categories	% Category 1	% Category 2	% Category 3
Physical Science	0%	83%	17%
Life Science	0%	83%	17%
Earth and Space Science	0%	83%	17%
Grade 5 Form 2	Cognitive Engagement by Domain		
Reporting Categories	% Category 1	% Category 2	% Category 3
Physical Science	17%	66%	17%
Life Science	0%	100%	0%
Earth and Space Science	17%	83%	0%
Grade 5 Form 3	Cognitive Engagement by Domain		
Reporting Categories	% Category 1	% Category 2	% Category 3
Physical Science	0%	100%	0%
Life Science	0%	100%	0%
Earth and Space Science	17%	66%	17%

Grade 8 Sample Test Events

The three West Virginia grade 8 sample test events analyzed were found to be fully or acceptably aligned based on the criteria used in this analysis. To fully meet state expectations, Form 2 was found to need three items revised or replaced and Form 3 was found to need one item revised or replaced to ensure no items with only Category 1 interactions and at least one item per domain that included one or more interactions requiring Category 3 cognitive engagement. Panelists included qualitative feedback and suggestions for many items across forms that merit consideration.

Overall, however, for Form 1, all domains fully met all alignment criteria. For Form 2, the Physical Science domain and Life Science domains weakly met Consistency of Cognitive Engagement and fully met all other criteria, and the Earth and Space Science domain fully met all alignment criteria. For Form 3, the Physical Science domain weakly met Consistency of Cognitive Engagement. Otherwise, all domains fully met all criteria.

Table 5. Distribution of Items by Category of Cognitive Engagement, West Virginia Grade 8 Forms 1-3

Grade 8 Form 1	Cognitive Engagement by Domain		
Reporting Categories	% Category 1	% Category 2	% Category 3
Physical Science	0%	50%	50%
Life Science	0%	83%	17%
Earth and Space Science	0%	67%	33%
Grade 8 Form 2	Cognitive Engagement by Domain		
Reporting Categories	% Category 1	% Category 2	% Category 3
Physical Science	33%	50%	17%
Life Science	0%	100%	0%
Earth and Space Science	0%	50%	50%
Grade 8 Form 3	Cognitive Engagement by Domain		
Reporting Categories	% Category 1	% Category 2	% Category 3
Physical Science	17%	83%	0%
Life Science	0%	83%	17%
Earth and Space Science	0%	67%	33%

**Findings: Overall West Virginia
General Summative Assessment (WVGSA) for Science
Program Capacity to Generate Aligned Assessments**

The West Virginia General Summative Assessment (WVGSA) for science was found to have the capacity to generate test forms for grades 5 and 8 that were all fully or acceptably aligned with the West Virginia Next Generation Content Standards and Objectives for Science (equivalent to the NGSS). This section includes findings related to the overall capacity for alignment of the WVGSA in terms of all nine criteria agreed upon by participating states to be used to evaluate alignment of the assessments with NGSS. The overall findings reported are grounded in a synthesis of the item-level, test-event-level, and item-bank-level findings.

Results from the item-level analysis of the overall item bank show that nearly all items met expectations for Use of Phenomenon, Dimensionality (Structure of Knowledge), required appropriate cognitive engagement, consistent with the expectations of the PEs (Consistency of Cognitive Engagement), and were associated with scoring assertions that reasonably reflected both the PEs and the actual student interactions. A total of 19 items from the Shared Science Item Bank for elementary and middle school were flagged for revision or removal due to one or more unmet expectations. West Virginia's item bank included 17 of these items: 12 from the grade 5 item bank and five from the grade 8 item bank (see Appendix C of Common Report). As described in the common alignment report, while these items should be reviewed for ongoing improvement, any disruption to alignment caused by these few items would be minimal. Of the items flagged, all but one is a stand-alone item and nearly all are associated with just one or two scoring assertions. Because of the relatively limited interactions, a single stand-alone item contributes proportionately minimally to a student's score. While CAI and states may wish to address these items in ongoing item bank maintenance, these few items were not considered a significant threat to the alignment of test events.

Based on the data from all administered test events and as reported in the WVGSA 2018-2019 Technical Report, essentially all test events for grades 5 and 8 met blueprint specifications. There were only very few exceptions (<0.0001% of tests; four students across both grades), with very minor blueprint violations that occurred under specific conditions, such as when a student took the same grade-level test two years in a row.

Study results show that the West Virginia WVGSA for science for grades 5 and 8 had the capacity for essentially all test events to fully or acceptably meet all of West Virginia's alignment expectations as defined by the criteria used in this study. This finding is based on a consideration of the West Virginia item bank, assessment blueprints, sample test events, and aggregate data from all administered test events, in the context of the greater item-bank-level findings.

Findings are reported for each of the nine alignment criteria used in this analysis, based on the item-level analysis of the overall item bank as applied to the West Virginia subset of items as well as an item-level analysis of three West Virginia sample test events from each of grades 5 and 8. Additional detail about each criterion is provided in the common alignment report and is not repeated here.

Criterion 1. Use of Phenomena

Based on the item-level analysis of the overall item bank, with the exception of four grade 5 items and three grade 8 items, all items within the West Virginia item bank were considered by a majority of panelists to meet states' expectations for being phenomenon-based. Therefore, it can be expected that all test events that use items from the West Virginia item bank will meet the expectation that at least 90% of items and item clusters require students to engage multiple dimensions of the standards ("use science") to make sense of phenomena.

Criterion 2. Categorical Concurrence

For a particular test event, this criterion is judged by determining whether the assessment included items targeting PEs from each reporting category. Grounded in calculations based on a procedure developed by Subkoviak (1988), it is typically assumed that an assessment would have to have at least six items for measuring content from a reporting category for a minimum acceptable level of Categorical Concurrence to exist between the domain and the assessment (Webb, 1999). The number of items (six) is based on estimating the number of items that could produce a reasonably reliable score for estimating students' mastery of content on that subscale. Of course, many factors must be considered in determining what a reasonable number is, including the reliability of the subscale, the mean score, and cutoff score for determining mastery.

A cutoff of six items per reporting category was consistent with West Virginia expectations and was used in this analysis. For all grades, the West Virginia blueprint specifies six items for each reporting category. Based on the results of the item-level analysis of all items in the West Virginia item bank (available as of July 2019), the internally coded standard (PE) is appropriate for all items/clusters, with the exception of a very small proportion items. Based on the aggregate data as reported in the WVGSA 2018-2019 Technical Report, all

administered test events met blueprint specifications at the domain level. Therefore, with consideration of the item-level analysis results and the information from aggregate data, it can be inferred that all test events for the WVGSA for Science for grades 5 and 8 met the criterion of Categorical Concurrence, with the possible exception of the forms that contained the small number of particular items flagged by reviewers as inadequate to address the PE. Note, however, that because item clusters contain multiple interactions, with an average of approximately 10 interactions, and because all but one of the flagged items were stand-alone items, it is expected that a student would still have sufficient opportunity to interact with the main categories of content even if a test form contained one of these flagged items.

Essentially all tests generated by the WVGSA for Science for grades 5 and 8 can be inferred to have met the criterion of Categorical Concurrence with the corresponding West Virginia Next Generation Content Standards and Objectives for Science based on a consideration of the results of the item-level analyses of the overall item bank, the West Virginia item bank, sample state test events, West Virginia blueprints, and aggregate data from all administered West Virginia test events.

3. Dimensionality (Structure of Knowledge)

Based on the item-level analysis of the overall item bank and disregarding the small proportion of items flagged for revision or removal, all items within the West Virginia item bank were considered by a majority of panelists to meet West Virginia's expectations for dimensionality. All item clusters were found to require students to demonstrate integrated engagement with the three dimensions of Science and Engineering Practices (SEPs), Disciplinary Core Ideas (DCIs), and Crosscutting Concepts (CCCs) specified in the targeted PE. All stand-alone items were found to require students to demonstrate integrated engagement with two or three of the dimensions specified in the targeted PE. Therefore, all tests generated by the WVGSA for Science for grades 5 and 8 can be inferred to have met the criterion of Dimensionality (Structure of Knowledge) as relates to the corresponding West Virginia Next Generation Content Standards and Objectives for Science.

4. Consistency of Cognitive Engagement

Consistency of Cognitive Engagement between content standards and an assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the corresponding standards. For consistency to exist between the assessment and the reporting categories, as judged in this analysis, two conditions apply. First, no items or item clusters should require only

Category 1 cognitive engagement. While it is acceptable for some interactions within an item or item cluster to be Category 1, successful completion of an item or item cluster cannot require only Category 1 work, per West Virginia expectations and consistent with the intent of the NGSS. Second, the proportion of items and item clusters with Category 2 and Category 3 opportunities should reflect the proportion of Category 2 and Category 3 expectations as represented within the Performance Expectations. Category 4 expectations, which are complex tasks that require extended time (such as the “sustained investigations” expected by the Framework) are not expected to be appropriately or authentically assessed in an on-demand context. All of the items and item clusters in the Shared Science Assessment Item Bank, therefore, are expected to provide opportunities for cognitive engagement within Category 2 and Category 3. Approximately 98% of items in the West Virginia item bank were considered Category 2 or 3 (**Table 6**). Consequently, it can be expected that all WVGSA for science for grades 5 and 8 generally have the capacity to meet the criterion of Consistency of Cognitive Engagement for all domains. This is supported by the findings from the West Virginia test events analyzed. For each of the grade 5 and 8 sample test events analyzed, all domains for both grades fully or weakly met this criterion, with the exception of the Physical Science domain on one grade 8 test form. Because the blueprint does not specify any distribution for the complexity of items, some variation in distribution of Category 2 and 3 is expected between and among forms. Nearly all instances of Category 3 cognitive engagement were found within item clusters. Therefore, the variation in distribution of Category 2 and 3 cognitive engagement between and among test forms will depend almost entirely on the particular item clusters assigned for each domain on a test event. On each of the sample test events for West Virginia, between one and five of the six item clusters included at least one opportunity for Category 3 cognitive engagement.

Table 6. Performance Expectations by Category of Engagement Compared with Category of Engagement Distribution within the West Virginia Item Bank

NGSS Grade Bands	Total Number of PEs	Category of Cognitive Engagement	Number of PEs by Level	Standards % Category of Engagement Distribution by Level	Items % Category of Engagement Distribution by level
Grades 3-5	42	1	0	0	2
		2	32	76	78
		3	8	19	20
		4	2	5	0
Grades 6-8	55	1	0	0	2
		2	39	71	58
		3	13	24	40
		4	3	5	0

5. Range of Knowledge Correspondence (Individual)

An appropriate Range of Knowledge for individual test events is defined by the West Virginia assessment blueprints and grounded in stakeholder deliberation and examination of multiple factors. Based on the item-level analysis of the overall item bank and disregarding the small proportion of items flagged for revision or removal, the items within the West Virginia item bank were considered by a majority of panelists to meet West Virginia's expectations at the item level. Based on the aggregate data as documented in the WVGSA 2018-2019 Technical Report, essentially all test events met blueprint specifications. Therefore, with consideration of the item-level analysis results and the information from aggregate data, and supported by the analysis of sample test events, it can be inferred that all test events for the WVGSA for Science for grades 5 and 8 met the criterion of Range of Knowledge Correspondence.

6. Range of Knowledge Correspondence (Population)

All items included in the July 2019 item-level analysis that were identified as West Virginia items as represented in Appendix C of the WVGSA 2018-2019 Technical Report Volume 1 are shown in **Table 2** by grade and domain. To identify the subset of items used in the West Virginia item bank, the item Master Identification Numbers from Appendix C were compared against the Master Identification Numbers from the items analyzed in the item-level alignment study. Any items that were not identified as West Virginia items in Appendix C were then eliminated from the item pool used to judge the extent to which the West Virginia item bank met state expectations to have the potential to address at least 90% of PEs.

The West Virginia science item bank met the state's expectations to contain items (stand-alone or cluster) addressing at least 90% of PEs within each domain for the elementary and middle school grade bands, and within the Life Science domain for high school (**Table 7**).

Table 7. WVGSA Science Assessment Item Bank Range of Knowledge Correspondence by Grade Band and Domain

Domain	Total Number of PEs	Number of PEs Not Represented	Percentage of PEs Targeted by Items Within Item Bank
Grade 3-5			
PS	17	1	100%
LS	12	1	100%
ESS	13	1	92%
Grade 6-8			
PS	19	1	95%
LS	21	2	90%
ESS	15	1	100%

Item exposure data as documented in the WVGSA 2018-2019 Technical Report show that the vast majority of science items (~98% or more) were administered on between 1% and 40% of test events. Item exposure was calculated by dividing the total number of test events in which an item appeared by the total number of tests administered. This finding provides additional evidence that the assessment fully met the criterion of Range of Knowledge (Population) for all reporting categories for grades 5 and 8.

7. Balance of Representation

Based on the item-level analysis of the overall item bank and disregarding the small proportion of items flagged for revision or removal, the items within the West Virginia item bank were considered by a majority of panelists to meet state expectations at the item level. Based on the aggregate data as documented in the WVGSA 2018-2019 Technical Report, all test events met blueprint specifications. Therefore, with consideration of item-level analysis results, information from aggregate data from all administered West Virginia science assessments, and supported by the analysis of sample test events, it can be inferred that all test events for the WVGSA for science for grades 5 and 8 met the criterion of Balance of Representation.

8. Relationship of Scoring Assertions with Student Interactions

Based on the item-level analysis of the overall item bank, all but one item within the West Virginia item bank were considered by a majority of panelists to meet states' expectations that, in aggregate, the scoring assertions for an item/item cluster appropriately represented the inferences about student knowledge, skills, and abilities that can be made based on successful interactions with an item/cluster. Therefore, it can be expected that all test events that use items from the West Virginia item bank will also meet the state's expectations as relates to the overall relationship between scoring assertions and student interactions.

9. Relationship of Scoring Assertions with PEs

Based on the item-level analysis of the overall item bank, with the exception of one grade 5 item and two grade 8 items, all items within the West Virginia item bank were considered by a majority of panelists to meet states' expectations that, in aggregate, the scoring assertions for an item appropriately represented the three-dimensional expectations within the targeted PE (at least two of the three dimensions for stand-alone items). Therefore, there is a strong likelihood that all test events that use items from the West Virginia item bank will also meet the state's expectations as relates to the overall relationship between scoring assertions and PEs.

It is important that the scoring assertions appropriately represent both the grade-level content standards and the student interactions on the assessment as the scoring assertions were used to define the performance standards for the assessment. In the test development process, CAI facilitated panelists in the use of an ordered-item process based on ordered sets of scoring assertions. West Virginia defines four Achievement Levels: Does Not Meet, Partially Meets, Meets, and Exceeds. Achievement-level descriptors (ALDs) define the content-area knowledge, skills, and processes that students at each performance level can demonstrate.

Conclusion

This report summarized the results of a content alignment analysis of the West Virginia General Summative Assessment (WVGSA) for science for grades 5 and 8 with West Virginia Next Generation Content Standards and Objectives for Science (equivalent to the Next Generation Science Standards) as pertains to fulfilling requirements as stated in Federal statute. The West Virginia assessment used a particular state-vetted subset of items that were part of a Shared Science Assessment Item Bank. The item bank is managed by Cambium Assessment (CAI; formerly known as American Institutes for Research - AIR) and is shared by multiple states.

An item- and item-cluster-level content alignment analysis was completed in July 2019 to evaluate the alignment of the overall Shared Science Assessment Item Bank with the Next Generation Science Standards' (NGSS) Performance Expectations (PEs). A common report describes the overall results of the alignment analysis of the item bank, applicable to all states.

Although study results suggest that the Shared Science Assessment Item Bank has strong capacity, overall, to generate aligned test forms, state-specific evidence is also necessary for a judgement of the capacity to generate aligned test forms for a particular state context. The overall results of the item-level content analyses were applied to West Virginia's assessment framework and statistical documentation, in the context of the state-specific item subset, to make some inference about the capacity for alignment of all test events generated by the WVGSA for science for grades 5 and 8 with corresponding West Virginia Next Generation Content Standards and Objectives for Science. Results from the item-level analyses were mapped onto the item sets from three West Virginia sample test events for each of grades 5 and 8 to provide additional state-specific test-event-level alignment information.

The WVGSA for science grades 5 and 8 was found to be acceptably aligned with the corresponding state standards for science, meaning that it had the overall capacity to generate test forms that were all fully or acceptably aligned with the West Virginia Next Generation Content Standards and Objectives for Science (equivalent to the NGSS). This finding includes consideration of the results of the item-level analyses of the overall item bank, the West Virginia item bank, sample West Virginia test events, West Virginia blueprints, and aggregate data from all administered West Virginia test events.

The evidence to support this finding includes:

- The West Virginia test blueprint identified the state's intended sampling across reporting categories (as relates to Categorical Concurrence, Range of Knowledge (breadth) for individual test events, and Balance of Representation (emphasis)).
- Aggregate data from all administered test events within the state of West Virginia showed that the blueprints and item selection algorithm yielded test forms as expected.
- Overall, the items within the West Virginia item bank met the state's expectations (as relates to Use of Phenomena, Dimensionality, Consistency of Cognitive Engagement, and relationships with scoring assertions).
- The West Virginia item bank was found to fully meet the state's expectations for Range of Knowledge across the tested student population for all grades and domains.
- An analysis of three sample test events from each of grades 5 and 8 found that all test events were fully or acceptably aligned, based on the criteria agreed upon by states and used in this analysis.

At the test event level, some variation in the Consistency of Cognitive Engagement was expected between and among test forms because the blueprint did not specify any distribution for the complexity of items. Nearly all instances of Category 3 cognitive engagement were found within item clusters. Therefore, the variation in distribution of Category 2 and 3 cognitive engagement between and among test forms will depend almost entirely on the particular item clusters assigned for each domain on a test event. On each of the sample test events for West Virginia, between one and five of the six item clusters included at least one opportunity for Category 3 cognitive engagement. The overall distribution of complexity of items in the item bank was found to be appropriate (Category 2 and 3) in relation to the distribution of complexity in the Performance Expectations. As such, Consistency of Cognitive Engagement of the assessment with standards can be expected across the tested student population.

Information about the item-bank-level and test-event-level distribution of WVGSA items by Category of Cognitive Engagement is provided in this report. If West Virginia wishes to have greater consistency in the distribution of items by Category of Cognitive Engagement between and among test forms, adjustments would need to be made to the item bank and/or to the test blueprints. Adjustments could also help ensure that all test events include at least one item per domain that requires Category 3 cognitive engagement. West Virginia may additionally wish to consider revision or removal of the specific items flagged in the item-level analysis as well as consider panelist feedback to support ongoing maintenance of and improvement to the item bank.

References

- Achieve, Inc.; Stanford/SCALE (2019) Reconceptualizing Alignment for NGSS Assessments. NARST Annual Meeting. Symposium paper available (01/2020) at https://snapgse.stanford.edu/sites/g/files/sbiybj10126/f/narst_symposium_2019_3.31.19.pdf
- Badrinarayan, A., Christopherson, S. and Harris, C. (2017) The Alignment Challenge. Reidy Interactive Lecture Series: Assessing Student Learning of the Next Generation Science Standards. The National Center for the Improvement of Educational Assessment
- Badrinarayan, A, Christopherson, S., Gong, B., McCrae, A. (2018) Developing a Common Language to Understand Content Complexity for Alignment Studies of the NGSS. National Conference on Student Assessment
- Badrinarayan, A., Christopherson, S., Davis-Becker, S., Everson, H., and Forte, E. (2019). Representing Cognitive Complexity in Test Design and Evaluation. ATP Innovations in Testing
- National Research Council. 2012. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press. <https://doi.org/10.17226/13165>.
- National Research Council. (2005). How Students Learn: History, Mathematics, and Science in the Classroom. Committee on How People Learn, A Targeted Report for Teachers, M.S. Donovan and J.D. Bransford, Editors. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- NGSS Lead States. 2013. Next Generation Science Standards: For States, By States. Washington, DC: The National Academies Press.
- NGSS Lead States. 2013. NGSS Release: How to Read the Next Generation Science Standards (NGSS). Washington, DC: The National Academies Press.
- Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25(1), 47-55.

- Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literary standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III*. Mahwah, NJ: Lawrence Erlbaum.
- Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education. Council of Chief State School Officers and National Institute for Mathematics Education Research Monograph No. 6. Madison: University of Wisconsin, Wisconsin Center for Education Research.
- Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison, WI: University of Wisconsin.
- Webb, N. L. (1999). Alignment of science and mathematics standards and assessments in four states. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 18. Madison, WI: University of Wisconsin.
- Webb, N. L. (2003). Alignment study in language arts, mathematics, science, and social studies of state standards and assessments in four states. Washington, D. C.: Council of Chief State School officers.
- Wise, S. L., Kingsbury, G. G., and Webb, N. L. (2015). Evaluating content alignment in computerized adaptive testing. *Educational Measurement: Issues and Practices*, 34, 4, pp. 41-48.
- WVGSA 2018-2019 Technical Report (2019). West Virginia Department of Education / American Institutes for Research.

The Impact of the Pandemic on West Virginia Students

Introduction

Since the onset of the pandemic, K-12 teachers and students have experienced unprecedented educational challenges with ongoing school closures due to the pandemic. The resulting disruptions in student learning led to significant concerns across states regarding student performance on the spring 2021 statewide educational assessments, with fears that scores would be compromised due to the lack of instructional coverage of content, differences in test modalities (remote versus onsite testing), and barriers to learning, such as a lack of internet access and varying levels of instructional and emotional support at home (Boyer, Dadey, & King, 2020). These fears have continued into 2022, as additional school closures continue to occur in some areas with periodic surges in COVID-19 infection. The Council of Chief State School Officers (CCSSO, 2020a) has made several suggestions for ways in which the potential impact of the pandemic on learning can be investigated, such as the comparison of item statistics to previous years' test administrations and extensions of regularly performed analyses such as descriptive statistics and test reliability.

The WVGSA test window opened on April 4, 2022. Regardless of the circumstances, CAI psychometricians immediately begin to monitor item and overall assessment performance once the test becomes available to students. With the concerns over the changes in learning since the onset of the pandemic, in addition to their usual monitoring procedures, the CAI psychometric team began to incorporate additional analyses into their processes to assist the West Virginia Department of Education (WVDE) in the identification of any changes in performance since the pre-pandemic assessment of Spring of 2019. The present investigation seeks to continue this evaluation to see if there have been any changes in student performance since the original study was conducted.

This report could provide WVDE with additional information on their ongoing efforts in remediation and support in the presence of any student gaps in learning due to the pandemic. Prior to analyses, the samples were checked with respect to gender, ethnicity, and special group representation to ensure they were consistent with the overall student population (Tables 1 - 5).

Specific research questions were as follows:

- With the changes in instruction introduced by the pandemic, did student academic performance change between 2021 and 2022 as evidenced by changes in overall scale scores, performance levels, correlational relationships between variables, and residuals?
- Have the test reliability coefficients and item parameters obtained during the current test administration changed since the spring 2021 administration?

Table 1: West Virginia Spring 2022 ELA Student Gender Distributions

Grade	Gender	N_2019	Percent_2019	N_2021	Percentage_2021	N	Percentage_2022	Difference
3	Female	9060	49%	8010	49%	8538	49%	0
3	Male	9499	51%	8326	51%	8988	51%	0
4	Female	9479	49%	8073	49%	8454	49%	0
4	Male	9736	51%	8543	51%	8869	51%	0
5	Female	9482	48%	8201	49%	8611	49%	0
5	Male	10073	52%	8559	51%	9072	51%	0
6	Female	9486	48%	8300	49%	8678	49%	0
6	Male	10116	52%	8569	51%	9019	51%	0
7	Female	9124	48%	8320	48%	8979	49%	1
7	Male	9714	52%	8931	52%	9263	51%	-1
8	Female	9095	48%	8267	48%	9012	48%	0
8	Male	9959	52%	8976	52%	9686	52%	0

Table 2: West Virginia Spring 2022 Math Student Gender Distributions

Grade	Gender	N_2019	Percent_2019	N_2021	Percentage_2021	N	Percentage_2022	Difference
3	Female	9058	49%	8020	49%	8548	49%	0
3	Male	9502	51%	8328	51%	8994	51%	0
4	Female	9473	49%	8073	49%	8459	49%	0
4	Male	9732	51%	8553	51%	8870	51%	0
5	Female	9482	48%	8221	49%	8632	49%	0
5	Male	10075	52%	8596	51%	9085	51%	0
6	Female	9479	48%	8339	49%	8682	49%	0
6	Male	10109	52%	8625	51%	9026	51%	0
7	Female	9129	48%	8366	48%	8994	49%	1
7	Male	9720	52%	8983	52%	9287	51%	-1
8	Female	9088	48%	8333	48%	9019	48%	0
8	Male	9968	52%	9035	52%	9699	52%	0

Table 3: West Virginia Spring 2022 ELA Student Ethnicity Distributions

Grade	Ethnicity	N_2019	Percent_2019	N_2021	Percent_2021	N_22	Percent_2022	Difference
3	Afr_American	714	4%	582	4%	685	4%	0
3	Amer_Indian	8	0%	8	0%	6	0%	0
3	Asian	110	1%	86	1%	102	1%	0
3	Hispanic	380	2%	309	2%	370	2%	0
3	Multi-ethn	750	4%	723	4%	800	5%	1
3	Pac_Islander	11	0%	7	0%	10	0%	0
3	White	16575	89%	14599	89%	15252	87%	-2
4	Afr_American	783	4%	614	4%	630	4%	0
4	Amer_Indian	22	0%	10	0%	8	0%	0
4	Asian	122	1%	107	1%	98	1%	0
4	Hispanic	381	2%	319	2%	370	2%	0
4	Multi-ethn	770	4%	697	4%	794	5%	1
4	Pac_Islander	7	0%	6	0%	8	0%	0
4	White	17130	89%	14852	89%	15129	87%	-2
5	Afr_American	732	4%	643	4%	649	4%	0
5	Amer_Indian	19	0%	9	0%	7	0%	0
5	Asian	129	1%	95	1%	113	1%	0
5	Hispanic	411	2%	355	2%	372	2%	0
5	Multi-ethn	753	4%	691	4%	804	5%	1
5	Pac_Islander	12	0%	8	0%	8	0%	0
5	White	17499	89%	14956	89%	15439	87%	-2
6	Afr_American	790	4%	675	4%	683	4%	0
6	Amer_Indian	11	0%	23	0%	13	0%	0
6	Asian	137	1%	92	1%	100	1%	0
6	Hispanic	389	2%	327	2%	410	2%	0
6	Multi-ethn	745	4%	622	4%	723	4%	0
6	Pac_Islander	9	0%	3	0%	9	0%	0
6	White	17519	89%	15126	90%	15282	86%	-4
7	Afr_American	825	4%	626	4%	783	4%	0
7	Amer_Indian	22	0%	19	0%	28	0%	0
7	Asian	120	1%	101	1%	117	1%	0
7	Hispanic	371	2%	362	2%	390	2%	0
7	Multi-ethn	610	3%	627	4%	696	4%	0
7	Pac_Islander	12	0%	7	0%	3	0%	0
7	White	16878	90%	15509	90%	15946	87%	-3
8	Afr_American	835	4%	636	4%	717	4%	0
8	Amer_Indian	16	0%	10	0%	20	0%	0
8	Asian	137	1%	124	1%	120	1%	0
8	Hispanic	385	2%	329	2%	440	2%	0
8	Multi-ethn	529	3%	609	4%	729	4%	0
8	Pac_Islander	8	0%	6	0%	9	0%	0
8	White	17144	90%	15529	90%	16371	88%	-2

Table 4: West Virginia Spring 2022 Math Student Ethnicity Distributions

Grade	Ethnicity	N_2019	Percent_2019	N_2021	Percent_2021	N_22	Percent_2022	Difference
3	Afr_American	714	4%	586	4%	686	4%	0
3	Amer_Indian	8	0%	8	0%	6	0%	0
3	Asian	111	1%	85	1%	102	1%	0
3	Hispanic	377	2%	305	2%	367	2%	0
3	Multi-ethn	751	4%	726	4%	802	5%	1
3	Pac_Islander	11	0%	7	0%	10	0%	0
3	White	16577	89%	14608	89%	15267	87%	-2
4	Afr_American	783	4%	615	4%	631	4%	0
4	Amer_Indian	22	0%	10	0%	8	0%	0
4	Asian	122	1%	107	1%	98	1%	0
4	Hispanic	377	2%	317	2%	359	2%	0
4	Multi-ethn	769	4%	698	4%	797	5%	1
4	Pac_Islander	7	0%	5	0%	8	0%	0
4	White	17125	89%	14863	89%	15140	87%	-2
5	Afr_American	734	4%	642	4%	651	4%	0
5	Amer_Indian	19	0%	9	0%	7	0%	0
5	Asian	130	1%	95	1%	113	1%	0
5	Hispanic	400	2%	352	2%	370	2%	0
5	Multi-ethn	754	4%	689	4%	806	5%	1
5	Pac_Islander	12	0%	8	0%	8	0%	0
5	White	17508	90%	15019	89%	15475	87%	-2
6	Afr_American	794	4%	676	4%	708	4%	0
6	Amer_Indian	11	0%	25	0%	13	0%	0
6	Asian	137	1%	92	1%	100	1%	0
6	Hispanic	379	2%	323	2%	402	2%	0
6	Multi-ethn	747	4%	628	4%	741	4%	0
6	Pac_Islander	9	0%	3	0%	9	0%	0
6	White	17509	89%	15216	90%	15433	87%	-3
7	Afr_American	825	4%	638	4%	788	4%	0
7	Amer_Indian	22	0%	19	0%	29	0%	0
7	Asian	120	1%	101	1%	118	1%	0
7	Hispanic	366	2%	352	2%	377	2%	0
7	Multi-ethn	611	3%	629	4%	698	4%	0
7	Pac_Islander	12	0%	8	0%	3	0%	0
7	White	16893	90%	15602	90%	15989	87%	-3
8	Afr_American	836	4%	645	4%	718	4%	0
8	Amer_Indian	16	0%	10	0%	19	0%	0
8	Asian	137	1%	125	1%	120	1%	0
8	Hispanic	376	2%	319	2%	419	2%	0
8	Multi-ethn	529	3%	614	4%	732	4%	0
8	Pac_Islander	8	0%	6	0%	9	0%	0
8	White	17154	90%	15649	90%	16407	88%	-2

Table 5: West Virginia Spring 2022 Student LEP Distributions

Subject	Grade	EL	N_2019	Percent_2019	N_2021	Percent_2021	N	Percent_2022	Difference
ELA	3	N	18402	99%	16233	99%	17058	97%	-2
ELA	3	Y	157	1%	119	1%	185	1%	0
ELA	4	N	19051	99%	16485	99%	16927	98%	-1
ELA	4	Y	164	1%	136	1%	131	1%	0
ELA	5	N	19407	99%	16674	99%	17272	98%	-1
ELA	5	Y	148	1%	94	1%	131	1%	0
ELA	6	N	19480	99%	16785	99%	17123	97%	-2
ELA	6	Y	122	1%	92	1%	100	1%	0
ELA	7	N	18731	99%	17159	99%	17843	98%	-1
ELA	7	Y	107	1%	108	1%	122	1%	0
ELA	8	N	18954	99%	17163	99%	18273	98%	-1
ELA	8	Y	100	1%	97	1%	133	1%	0
Math	3	N	18407	99%	16236	99%	17077	97%	-2
Math	3	Y	153	1%	115	1%	181	1%	0
Math	4	N	19046	99%	16506	99%	16941	98%	-1
Math	4	Y	159	1%	131	1%	121	1%	0
Math	5	N	19420	99%	16734	99%	17312	98%	-1
Math	5	Y	137	1%	90	1%	129	1%	0
Math	6	N	19475	99%	16890	99%	17318	98%	-1
Math	6	Y	113	1%	85	1%	91	1%	0
Math	7	N	18746	99%	17266	99%	17896	98%	-1
Math	7	Y	103	1%	93	1%	108	1%	0
Math	8	N	18965	100%	17292	99%	18315	98%	-1
Math	8	Y	91	0%	88	1%	109	1%	0

1. Changes in Achievement Levels

A critical indicator of student academic performance on the state’s content standards is the classification of student performance by achievement levels defined by cut scores established during a standard setting meeting conducted by CAI in 2018. Tables 6, 7, and 8 demonstrate the differences observed between spring 2019, spring 2021, and spring 2022 proficiency and performance levels.

Table 6: West Virginia ELA Student Achievement

Grade	AL	2018 Percentage	2019 Percentage	2021 Percentage	2022 Percentage	Impact
3	1	21%	25%	35%	34%	-1
3	2	32%	32%	31%	29%	-2
3	3	29%	26%	22%	21%	-1
3	4	18%	17%	12%	15%	3
4	1	26%	24%	33%	30%	-3
4	2	29%	28%	30%	27%	-3
4	3	24%	25%	21%	23%	2
4	4	20%	23%	16%	21%	5
5	1	27%	26%	32%	31%	-1
5	2	29%	27%	28%	28%	0
5	3	27%	26%	24%	24%	0
5	4	17%	21%	16%	16%	0
6	1	27%	25%	27%	27%	0
6	2	30%	31%	34%	31%	-3
6	3	29%	31%	28%	30%	2
6	4	14%	14%	11%	12%	1
7	1	26%	27%	30%	29%	-1
7	2	31%	31%	32%	31%	-1
7	3	29%	29%	26%	27%	1
7	4	15%	13%	12%	13%	1
8	1	27%	25%	27%	31%	4
8	2	32%	32%	30%	30%	0
8	3	28%	29%	28%	25%	-3
8	4	14%	15%	15%	14%	-1

Table 7: West Virginia Math Student Achievement Impact

Grade	AL	2018 Percentage	2019 Percentage	2021 Percentage	2022 Percentage	Impact
3	1	23%	21%	32%	27%	-5
3	2	29%	29%	29%	27%	-2
3	3	27%	26%	23%	25%	2
3	4	21%	25%	16%	21%	5
4	1	22%	20%	31%	26%	-5
4	2	33%	33%	35%	33%	-2
4	3	22%	22%	18%	20%	2
4	4	23%	25%	16%	21%	5
5	1	29%	26%	38%	34%	-4
5	2	31%	34%	33%	31%	-2
5	3	20%	20%	16%	18%	2
5	4	20%	20%	13%	17%	4
6	1	34%	34%	46%	42%	-4
6	2	33%	32%	33%	31%	-2
6	3	19%	19%	14%	16%	2
6	4	14%	15%	7%	11%	4
7	1	33%	34%	42%	41%	-1
7	2	31%	30%	32%	30%	-2
7	3	21%	19%	16%	17%	1
7	4	14%	17%	10%	12%	2
8	1	34%	33%	45%	43%	-2
8	2	34%	31%	31%	30%	-1
8	3	14%	15%	12%	12%	0
8	4	17%	21%	12%	15%	3

Table 8: West Virginia Student Proficiency Level Impact

Grade	Subject	2018_Proficient+	2019_Proficient+	2021_Proficient+	2022_Proficient+	Impact
3	ELA	47%	43%	33%	36%	3
4	ELA	44%	48%	37%	44%	7
5	ELA	44%	47%	40%	41%	1
6	ELA	43%	45%	39%	42%	3
7	ELA	44%	42%	38%	40%	2
8	ELA	42%	44%	43%	39%	-4
3	Math	48%	51%	38%	46%	8
4	Math	45%	47%	34%	41%	7
5	Math	40%	40%	28%	34%	6
6	Math	33%	34%	21%	27%	6
7	Math	35%	36%	26%	29%	3
8	Math	31%	36%	24%	27%	3

2. Changes in Scale Scores

The mean scale scores provide a point estimate that can be compared to previous years' mean scores, while the standard deviation provides us with a measure of the overall spread of scores from the mean. For example, if scores have potentially been impacted by the pandemic, we might expect to see some significant downward shifts in the mean scores and/or larger standard deviations when compared to the 2019 scores (although additional analyses would be required to determine if such differences were statistically significant).

Variability between different test administrations is always expected regardless of the circumstances. Therefore, in an effort to reduce some of this variability, it was also useful to compare spring 2022 scores to the overall mean scale scores of spring 2019 and spring 2021 (Table 9).

Table 9: West Virginia Spring 2022 Scale Score Impact

Grade	Subject	2018_N	2018_Mean	2018_SD	2019_N	2019_Mean	2019_SD	2021_N	2021_Mean	2021_SD	2022_N	2022_Mean	2022_SD	Impact
3	ELA	19343	580	39.10	18559	577	40.03	16352	566	41	17526	568	44	2
4	ELA	19751	591	43.84	19215	594	46.02	16621	583	46	17323	588	48	5
5	ELA	19874	612	44.30	19555	617	46.03	16768	608	45	17683	609	46	1
6	ELA	19085	627	48.21	19602	629	47.89	16877	624	46	17697	625	48	1
7	ELA	19280	633	49.60	18838	631	48.50	17267	627	49	18242	628	50	1
8	ELA	19374	644	48.84	19054	646	49.49	17260	644	51	18698	639	53	-5
3	Math	19409	422	34.18	18560	425	33.74	16351	414	36	17542	420	36	6
4	Math	19790	449	41.72	19205	450	41.27	16637	438	42	17329	445	44	7
5	Math	19934	472	52.07	19557	474	48.24	16824	459	49	17717	465	51	6
6	Math	19125	493	55.01	19588	492	57.06	16975	477	53	17708	482	57	5
7	Math	19324	523	59.95	18849	523	64.42	17359	510	60	18281	513	63	3
8	Math	19444	552	72.24	19056	555	76.90	17380	534	73	18718	537	77	3

3. Changes in Test Reliability

Test reliability provides information regarding how consistently a test measures the construct of interest. Theoretically, if a student was tested twice using an assessment with a high reliability coefficient, we would expect that the student would receive a similar score reflective of their true ability level with both test administrations. However, if the same student tested the second time while feeling extremely ill, it is unlikely that a similar score would be produced due to the variation in measurement introduced by the illness (Haertel, 2006). Such variations are referred to as measurement error, and when a high degree of error is introduced from external influences, test reliability coefficients can be negatively impacted.

If a significant decrease in reliability is observed, this might suggest that additional measurement error may have been introduced due to the changes in either student learning or the testing conditions (Table 10). Because there is no set form in adaptive testing, marginal reliability was computed for the scale scores, taking into account the varying measurement errors across the ability range. Marginal reliability is a measure of the overall reliability of an assessment based on the average conditional standard error of measurement, estimated at different points on the ability scale, for all students.

Table 10: West Virginia Spring 2022 Test Reliability Impact

Grade	Subject	2018_N	2018_Reliability	2019_N	2019_Reliability	2021_N	2021_Reliability	2022_N	2022_Reliability	Impact
3	ELA	19343	0.89	18559	0.89	16352	0.89	17526	0.89	0.00
4	ELA	19751	0.88	19215	0.88	16621	0.87	17323	0.89	0.02
5	ELA	19874	0.89	19555	0.90	16768	0.89	17683	0.90	0.01
6	ELA	19085	0.90	19602	0.89	16877	0.89	17697	0.89	0.00
7	ELA	19280	0.88	18838	0.90	17267	0.90	18242	0.90	0.00
8	ELA	19374	0.90	19054	0.91	17260	0.91	18698	0.91	0.00
3	Math	19409	0.92	18560	0.92	16351	0.91	17542	0.91	0.00
4	Math	19790	0.92	19205	0.92	16637	0.91	17329	0.92	0.01
5	Math	19934	0.93	19557	0.92	16824	0.89	17717	0.90	0.01
6	Math	19125	0.92	19588	0.89	16975	0.88	17708	0.88	0.00
7	Math	19324	0.93	18849	0.87	17359	0.87	18281	0.88	0.01
8	Math	19444	0.91	19056	0.89	17380	0.88	18718	0.89	0.01

4. Changes in Item Statistics

Item statistics are regularly monitored by CAI psychometricians throughout the testing window. Item difficulty (p-value) is the average item scores across all students with lower numbers indicating higher difficulty levels. Significant increases in item difficulty would have suggested that additional investigations may be warranted (Tables 11-12). Items that presented a decrease in p-value greater than 0.1 were flagged, and the percentage of the flagged items within each content standard was computed. The items that were administered both in spring 2021 and 2022 (common items in the pool) to more than 100 students (items compared) were included in this p-value evaluation.

Table 11: West Virginia Spring 2021 - 2022 Item PValues ELA

Grade	ContentLevelID	N_common	N_Compared	N_Flagged	Prop_Flagged
3	IT 3.IKI ELA.3.17	9	9	2	22%
3	IT 3.KID ELA.3.5	18	17	1	6%
3	L 3.VAU ELA.3.40a	12	12	2	17%
3	L 3.CSE ELA.3.37f	3	2	1	50%
3	L 3.CSE ELA.3.36i	5	4	1	25%
3	L 3.CSE ELA.3.36d	3	2	1	50%
3	L 3.CSE ELA.3.36g	6	4	1	25%
3	LT 3.KID ELA.3.3	23	22	2	9%
3	IT 3.KID ELA.3.6	20	17	2	12%
3	IT 3.KID ELA.3.4	29	28	2	7%
3	IT 3.CS ELA.3.11	18	18	2	11%
3	SL 3.CaC ELA.3.32	2	2	1	50%
3	IT 3.IKI ELA.3.16	10	8	2	25%
3	IT 3.CS ELA.3.10	16	15	1	7%
3	LT 3.CS ELA.3.8	17	16	1	6%
3	LT 3.IKI ELA.3.14	12	12	1	8%
4	L 4.VAU ELA.4.39a	14	14	3	21%
4	LT 4.KID ELA.4.3	20	17	1	6%
4	LT 4.KID ELA.4.1	14	14	3	21%
4	LT 4.KID ELA.4.2	12	12	1	8%
4	IT 4.KID ELA.4.5	25	21	1	5%
4	IT 4.KID ELA.4.4	29	26	1	4%
4	IT 4.KID ELA.4.6	21	18	1	6%
5	SL 5.CaC ELA.5.31	6	5	1	20%
5	IT 5.CS ELA.5.10	19	17	1	6%
5	IT 5.IKI ELA.5.16	20	17	1	6%
6	LT 6.IKI ELA.6.14	15	13	1	8%
6	IT 6.CS ELA.6.11	34	34	1	3%
6	LT 6.KID ELA.6.3	21	19	1	5%
6	LT 6.CS ELA.6.8	23	21	1	5%
6	LT 6.KID ELA.6.2	23	21	1	5%
6	L 6.VAU ELA.6.39a	19	18	1	6%
6	IT 6.KID ELA.6.4	38	34	1	3%
7	LT 7.CS ELA.7.7	21	20	2	10%
7	LT 7.CS ELA.7.9	17	17	2	12%
7	LT 7.KID ELA.7.3	25	25	1	4%
7	SL 7.CaC ELA.7.31	6	4	1	25%
7	IT 7.KID ELA.7.4	24	23	1	4%
8	IT 8.CS ELA.8.11	21	18	1	6%
8	IT 8.CS ELA.8.12	18	18	1	6%

8	IT 8.KID ELA.8.4	25	25	1	4%
8	IT 8.CS ELA.8.10	17	16	1	6%
8	LT 8.CS ELA.8.7	16	16	1	6%

Table 12: West Virginia Spring 2021 - 2022 Item PValues Math

Grade	ContentLevelID	N_common	N_Compared	N_Flagged	Prop_Flagged
3	MDG 3.G.c1 M.3.24	17	15	1	7%
3	MDG 3.G.c1 M.3.25	14	8	2	25%
3	MDG 3.MD.c1 M.3.16	17	16	2	12%
3	MDG 3.MD.c1 M.3.17	15	15	1	7%
3	MDG 3.MD.c2 M.3.18	18	18	3	17%
3	MDG 3.MD.c2 M.3.19	14	10	2	20%
3	MDG 3.MD.c3 M.3.22 M.3.22b	6	6	2	33%
3	NBTF 3.NBT.c1 M.3.11	34	34	4	12%
3	NBTF 3.NBT.c1 M.3.12	42	41	7	17%
3	NBTF 3.NF.c1 M.3.13	57	53	3	6%
3	NBTF 3.NF.c1 M.3.14 M.3.14b	45	22	1	5%
3	NBTF 3.NF.c1 M.3.15 M.3.15b	18	16	2	12%
3	OAT 3.OAT.c1 M.3.2	11	9	2	22%
3	OAT 3.OAT.c1 M.3.3	21	20	5	25%
4	MDG 4.G.c1 M.4.27	19	15	1	7%
4	MDG 4.G.c1 M.4.28	16	14	1	7%
4	MDG 4.MD.c1 M.4.20	16	15	1	7%
4	MDG 4.MD.c3 M.4.24	14	14	1	7%
4	NBTF 4.NBT.c2 M.4.9	30	30	1	3%
4	NBTF 4.NF.c1 M.4.12	28	21	1	5%
4	NBTF 4.NF.c2 M.4.14 M.4.14c	7	7	1	14%
4	NBTF 4.NF.c2 M.4.15 M.4.15b	12	12	1	8%
4	NBTF 4.NF.c3 M.4.17	24	17	1	6%
4	NBTF 4.NF.c3 M.4.18	25	25	1	4%
5	MDG 5.G.c1 M.5.23	17	15	4	27%
5	MDG 5.G.c1 M.5.24	20	19	1	5%
5	MDG 5.MD.c3 M.5.21	12	11	1	9%
5	MDG 5.MD.c3 M.5.22 M.5.22b	10	10	1	10%
5	NBTF 5.NBT.c1 M.5.4-5 M.5.5	18	8	2	25%
5	NBTF 5.NF.c1 M.5.11	28	28	1	4%
5	NBTF 5.NF.c2 M.5.14 M.5.14b	9	4	1	25%
5	NBTF 5.NF.c2 M.5.17 M.5.17b	9	6	1	17%
5	OAT 5.OAT.c1 M.5.1	34	33	1	3%
5	OAT 5.OAT.c2 M.5.3	18	18	1	6%
6	EE 6.EE.c1 M.6.12	19	18	1	6%
6	EE 6.EE.c1 M.6.14	26	21	3	14%
6	EE 6.EE.c1 M.6.15	19	14	1	7%
6	EE 6.EE.c2 M.6.17	18	14	1	7%
6	EE 6.EE.c2 M.6.19	21	18	2	11%
6	EE 6.EE.c3 M.6.20	31	20	1	5%
6	GSP 6.G.c1 M.6.22	20	20	1	5%
6	GSP 6.G.c1 M.6.23	15	15	4	27%
6	GSP 6.G.c1 M.6.24	15	10	1	10%
6	GSP 6.SP.c1 M.6.25	9	8	1	12%
6	RPNS 6.NS.c1 M.6.4	26	22	4	18%
6	RPNS 6.NS.c2 M.6.6	16	16	1	6%
6	RPNS 6.NS.c2 M.6.7	14	12	1	8%
6	RPNS 6.NS.c3 M.6.10 M.6.10d	6	3	1	33%
6	RPNS 6.NS.c3 M.6.8	10	10	2	20%
6	RPNS 6.NS.c3 M.6.9 M.6.9b	8	4	1	25%

6	RPNS 6.RP.c1 M.6.2	47	36	2	6%
7	EE 7.EE.c1 M.7.7	19	13	1	8%
7	EE 7.EE.c1 M.7.8	19	15	1	7%
7	EE 7.EE.c2 M.7.10 M.7.10a	12	7	1	14%
7	EE 7.EE.c2 M.7.10 M.7.10b	5	2	1	50%
7	G 7.G.c1 M.7.13	15	13	1	8%
7	G 7.G.c2 M.7.15	17	17	2	12%
7	G 7.G.c2 M.7.16	19	16	1	6%
7	RPNS 7.NS.c1 M.7.4 M.7.4c	10	7	2	29%
7	SP 7.SP.c1 M.7.18	12	9	1	11%
7	SP 7.SP.c2 M.7.21	10	5	1	20%
8	EENS 8.EE.c1 M.8.3	12	9	2	22%
8	EENS 8.EE.c1 M.8.5	12	10	1	10%
8	EENS 8.EE.c2 M.8.7	26	24	1	4%
8	EENS 8.EE.c2 M.8.8	20	12	1	8%
8	F 8.F.c1 M.8.11	20	16	2	12%
8	F 8.F.c2 M.8.15	22	17	1	6%
8	GSP 8.G.c1 M.8.16 M.8.16a	7	7	1	14%
8	GSP 8.G.c1 M.8.16 M.8.16b	3	3	2	67%
8	GSP 8.G.c1 M.8.19	12	12	1	8%
8	GSP 8.G.c1 M.8.20	16	16	1	6%
8	GSP 8.G.c2 M.8.21	12	8	1	12%
8	GSP 8.G.c2 M.8.22	17	14	1	7%
8	GSP 8.G.c3 M.8.24	26	12	1	8%
8	GSP 8.SP.c1 M.8.26	18	18	1	6%

5. Residual Analyses

In statistical models, a residual is the difference between the observed value and the expected value that the model predicts for that observation. If assessments fail to maintain the same level of score validity/stability between testing occasions, this may be manifested in the observed changes of the magnitudes and/or patterns of the residuals. For each item, a model-predicted (or expected) score was computed given the student's ability estimate and the item parameters known in the item bank using the IRT models adopted for the WVGSA. The item-level residuals were summed for a student (across all items) and for an aggregated unit (across all items and students in a grade). The aggregated residuals for pre- and post-pandemic administrations were compared at the test-level.

For dichotomous items, the 3PL IRT model was used to compute the expected score (1), and for polytomous items, the generalized partial credit model was used (2).

$$E(z_{ij}) = c_i + (1 - c_i) \frac{\exp(D a_i(\theta_j - b_i))}{1 + \exp(D a_i(\theta_j - b_i))}, \quad (1)$$

where, $E(z_{ij})$ is the expected score for item i for student j , c_i is the pseudo-guessing parameter, a_i is the item discrimination parameter, b_i is the item location (difficulty) parameter, and D is a constant fixed at 1.7, bringing the logistic into coincidence with the probit model. Student estimated ability is represented by θ_j .

$$E(z_{ij}) = \sum_{l=1}^{m_i} \frac{l \exp(\sum_{k=1}^l D a_i(\theta_j - b_{i,k}))}{1 + \sum_{l=1}^{m_i} \exp(\sum_{k=1}^l D a_i(\theta_j - b_{i,k}))}, \quad (2)$$

where, $b_{i,k}$ are the step parameter values with the maximum possible score m_i . For each item i the residual between observed and expected score for each student is defined as follows:

$$\delta_{ij} = z_{ij} - E(z_{ij}), \quad (3)$$

A positive δ_{ij} indicates that student j obtains a higher score on item i than the expected score given his or her ability and item parameters while a negative δ_{ij} indicates that student j performs lower than expected on item i .

The residuals (δ_{ij}) are summed across all items that student j was given and the sum of the residuals is divided by the total points possible for the test form student j took. Residual at test level for each individual is defined as (4):

$$\delta_j = \frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n (\delta_{ij}), \quad (4)$$

where m_i is the points possible for item i and n is the total number of items in the test form.

The residual analyses were conducted for spring 2019 WVGSA data to establish the reference values against which spring 2021 results were compared.

To investigate the changes in residuals at different ability levels, students were grouped into ten groups based on their estimated θ value. Within the lowest and highest theta boundaries for each grade, theta scales were divided into 10 bins. The bins with a lower number indicate a lower ability group and the groups with a higher number indicate a higher ability group.

The test-level residuals (4) were averaged across all students within a bin and the standard error (se) was computed.

$$\delta_g = \frac{1}{n_g} \sum_{j=1}^{n_g} (\delta_j), \quad (5)$$

$$se(\delta_g) = sd(\delta_g) / \sqrt{n_g}, \quad (6)$$

where g is the ability group, n_g is the number of students in group g , and the $sd(\delta_g)$ is the standard deviation of the test-level residual within g .

Note that there were no or few students in the higher bins for some tests. In the results section, the statistics calculated for those small groups were retained in the tables and plots to show the student's ability level distribution. However, the large standard error (6) associated with those small groups should be considered and caution is needed when interpreting the meaning of those values.

Figures 1 and 2 show the mean residual plot for ELA and mathematics, respectively. The means of residuals in each ability group (5) are marked by grade. As can be seen in the residual plots, there are significant outliers in some of the higher ability groups. However, Table 13 and 14 show that all of the outlier values are based on a very small number of students. The mean (M), standard deviation (SD) and the number of students (N) are included in these tables, and the lower and upper limit of the 95% confidence interval (CI) for the spring 2022 mean residuals are presented in the tables to assist with the interpretation of those values.

Table 13: ELA Residuals

Grd	Bin	N19	M19	SD19	N21	M21	SD21	N22	M22	SD22	CI_LL	CI_UL
3	1	60	0.0076	0.0412	59	0.0104	0.0405	93	0.0049	0.0440	-0.0042	0.0140
3	2	183	0.0069	0.0346	272	0.0091	0.0338	229	0.0062	0.0381	0.0012	0.0111
3	3	1245	0.0078	0.0347	1815	0.0112	0.0352	1108	0.0086	0.0351	0.0065	0.0107
3	4	3464	0.0063	0.0356	4020	0.0126	0.0362	2886	0.0110	0.0359	0.0097	0.0123
3	5	5490	-0.0037	0.0322	4702	0.0037	0.0341	3879	0.0072	0.0397	0.0060	0.0085
3	6	5333	-0.0019	0.0262	3806	0.0007	0.0298	4144	0.0036	0.0332	0.0026	0.0046
3	7	2444	0.0031	0.0249	1425	0.0032	0.0263	3401	0.0008	0.0272	-0.0001	0.0017
3	8	311	0.0064	0.0232	228	0.0123	0.0238	1516	0.0046	0.0270	0.0032	0.0059
3	9	21	0.0032	0.0281	9	-0.0015	0.0293	248	0.0067	0.0226	0.0038	0.0095
3	10	2	0.0046	0.0032	0	0.0000	0.0000	17	0.0034	0.0231	-0.0085	0.0152
4	1	150	0.0139	0.0355	139	0.0143	0.0357	140	0.0148	0.0394	0.0082	0.0214
4	2	408	0.0074	0.0303	484	0.0104	0.0321	483	0.0123	0.0351	0.0092	0.0155
4	3	1627	0.0035	0.0297	2025	0.0090	0.0312	1938	0.0073	0.0340	0.0058	0.0088
4	4	3766	-0.0012	0.0283	4241	0.0042	0.0309	3837	0.0001	0.0312	-0.0009	0.0011
4	5	5623	-0.0045	0.0274	4809	0.0027	0.0307	4780	-0.0024	0.0270	-0.0031	-0.0016
4	6	5062	-0.0022	0.0256	3396	-0.0007	0.0307	4074	-0.0024	0.0247	-0.0031	-0.0016
4	7	2204	0.0029	0.0244	1332	0.0025	0.0285	1715	-0.0008	0.0229	-0.0018	0.0003
4	8	349	0.0090	0.0243	177	0.0045	0.0274	324	0.0030	0.0217	0.0007	0.0054
4	9	23	0.0109	0.0258	13	0.0137	0.0249	24	0.0046	0.0251	-0.0060	0.0152
4	10	2	0.0250	0.0415	0	0.0000	0.0000	2	0.0240	0.0028	-0.0014	0.0494
5	1	91	0.0064	0.0347	73	0.0087	0.0349	62	0.0114	0.0347	0.0026	0.0202
5	2	345	0.0080	0.0297	392	0.0117	0.0309	216	0.0130	0.0315	0.0088	0.0172
5	3	1547	0.0026	0.0292	1790	0.0097	0.0308	997	0.0129	0.0326	0.0109	0.0149
5	4	3631	-0.0032	0.0305	3773	0.0017	0.0314	2362	0.0140	0.0331	0.0127	0.0154
5	5	5743	-0.0040	0.0283	4925	-0.0052	0.0293	3558	0.0035	0.0357	0.0023	0.0046
5	6	4993	-0.0012	0.0243	3993	-0.0034	0.0274	4409	-0.0007	0.0322	-0.0017	0.0002
5	7	2659	0.0037	0.0221	1590	-0.0009	0.0243	3689	0.0013	0.0284	0.0004	0.0022
5	8	509	0.0058	0.0221	216	0.0067	0.0215	1854	0.0042	0.0256	0.0030	0.0054
5	9	31	0.0031	0.0213	8	-0.0022	0.0290	484	0.0059	0.0217	0.0040	0.0079
5	10	4	0.0098	0.0179	0	0.0000	0.0000	44	0.0115	0.0157	0.0067	0.0162
6	1	166	0.0005	0.0344	115	0.0011	0.0319	138	0.0050	0.0350	-0.0009	0.0109
6	2	452	0.0024	0.0311	427	0.0042	0.0308	384	0.0063	0.0330	0.0030	0.0096
6	3	1837	-0.0002	0.0330	1792	0.0061	0.0313	1546	0.0020	0.0311	0.0005	0.0036
6	4	3723	-0.0006	0.0332	3619	0.0021	0.0345	3122	-0.0007	0.0326	-0.0018	0.0005
6	5	5532	-0.0080	0.0350	5075	-0.0058	0.0336	4577	-0.0103	0.0344	-0.0113	-0.0093
6	6	5205	-0.0065	0.0283	4046	-0.0047	0.0278	4708	-0.0111	0.0299	-0.0120	-0.0103
6	7	2239	0.0017	0.0277	1563	0.0004	0.0283	2519	-0.0034	0.0280	-0.0045	-0.0023
6	8	416	0.0008	0.0259	221	0.0031	0.0267	617	0.0015	0.0276	-0.0007	0.0037
6	9	25	0.0058	0.0193	9	-0.0104	0.0227	69	0.0026	0.0241	-0.0032	0.0083
6	10	3	-0.0120	0.0036	2	-0.0062	0.0221	6	-0.0017	0.0271	-0.0301	0.0268
7	1	121	0.0127	0.0335	126	0.0092	0.0331	136	0.0032	0.0315	-0.0022	0.0085
7	2	472	0.0101	0.0341	576	0.0091	0.0327	510	0.0075	0.0310	0.0048	0.0101
7	3	1941	0.0091	0.0351	2001	0.0101	0.0338	1882	0.0086	0.0325	0.0071	0.0100
7	4	3902	0.0062	0.0424	3796	0.0044	0.0397	3347	0.0053	0.0366	0.0040	0.0065
7	5	5545	0.0011	0.0365	5060	0.0002	0.0325	4940	-0.0028	0.0345	-0.0038	-0.0018
7	6	4729	0.0111	0.0303	3813	0.0015	0.0266	4579	0.0005	0.0272	-0.0003	0.0013
7	7	1824	0.0127	0.0303	1663	0.0030	0.0257	2311	0.0003	0.0242	-0.0006	0.0013
7	8	272	0.0097	0.0276	210	0.0094	0.0293	470	0.0031	0.0253	0.0008	0.0054
7	9	27	0.0025	0.0226	6	0.0131	0.0196	50	-0.0087	0.0275	-0.0165	-0.0009
7	10	0	0.0000	0.0000	0	0.0000	0.0000	5	-0.0186	0.0105	-0.0317	-0.0055

8	1	46	0.0081	0.0275	64	0.0150	0.0307	77	0.0158	0.0334	0.0082	0.0234
8	2	246	0.0104	0.0303	272	0.0062	0.0326	255	0.0158	0.0357	0.0114	0.0202
8	3	1630	0.0082	0.0338	1527	0.0066	0.0333	1254	0.0114	0.0336	0.0095	0.0132
8	4	3667	0.0010	0.0375	3493	-0.0034	0.0346	2914	0.0057	0.0347	0.0044	0.0070
8	5	5844	-0.0051	0.0325	5104	-0.0108	0.0307	4203	-0.0056	0.0326	-0.0065	-0.0046
8	6	5067	0.0013	0.0284	4496	-0.0055	0.0275	4602	-0.0032	0.0279	-0.0041	-0.0024
8	7	2145	0.0055	0.0263	1953	-0.0005	0.0255	3574	0.0003	0.0239	-0.0005	0.0011
8	8	371	0.0023	0.0258	315	0.0003	0.0228	1498	0.0023	0.0224	0.0011	0.0034
8	9	32	-0.0071	0.0121	18	-0.0118	0.0256	279	0.0022	0.0212	-0.0003	0.0047
8	10	4	-0.0042	0.0085	1	0.0091	0.0000	31	-0.0034	0.0156	-0.0091	0.0023

Table 14: Math Residuals

Grd	Bin	N19	M19	SD19	N21	M21	SD21	N22	M22	SD22	CI_LL	CI_UL
3	1	263	-0.0037	0.0223	427	-0.0033	0.0214	362	-0.0020	0.0231	-0.0044	0.0004
3	2	352	0.0008	0.0202	588	0.0006	0.0197	462	0.0014	0.0213	-0.0005	0.0034
3	3	955	-0.0028	0.0218	1405	-0.0001	0.0217	1321	-0.0004	0.0232	-0.0016	0.0009
3	4	2801	-0.0030	0.0218	3330	-0.0047	0.0240	3035	-0.0024	0.0234	-0.0032	-0.0015
3	5	5250	-0.0021	0.0212	4716	-0.0018	0.0223	4732	-0.0025	0.0224	-0.0032	-0.0019
3	6	5078	0.0000	0.0200	3764	-0.0013	0.0220	4572	-0.0021	0.0220	-0.0027	-0.0015
3	7	2931	0.0023	0.0195	1651	0.0017	0.0210	2314	0.0014	0.0214	0.0006	0.0023
3	8	777	0.0041	0.0183	398	0.0035	0.0196	612	0.0039	0.0198	0.0023	0.0054
3	9	123	0.0074	0.0158	50	0.0075	0.0152	115	0.0040	0.0165	0.0010	0.0071
3	10	28	0.0082	0.0147	15	0.0093	0.0180	15	0.0091	0.0206	-0.0024	0.0205
4	1	359	-0.0045	0.0207	457	-0.0061	0.0158	435	-0.0062	0.0175	-0.0079	-0.0046
4	2	425	0.0002	0.0188	608	-0.0013	0.0164	562	-0.0008	0.0194	-0.0024	0.0008
4	3	1141	-0.0021	0.0204	1729	-0.0060	0.0218	1374	-0.0047	0.0215	-0.0058	-0.0036
4	4	3250	-0.0057	0.0210	3680	-0.0073	0.0230	3350	-0.0062	0.0227	-0.0070	-0.0055
4	5	5886	-0.0046	0.0223	5209	-0.0050	0.0243	5192	-0.0027	0.0236	-0.0033	-0.0020
4	6	5074	-0.0013	0.0227	3305	-0.0007	0.0243	3950	0.0003	0.0239	-0.0005	0.0010
4	7	2404	0.0000	0.0217	1299	0.0000	0.0226	1928	0.0004	0.0225	-0.0006	0.0014
4	8	565	0.0037	0.0203	284	0.0026	0.0214	457	0.0047	0.0225	0.0026	0.0068
4	9	84	0.0030	0.0194	45	0.0052	0.0192	78	0.0047	0.0196	0.0003	0.0092
4	10	17	0.0072	0.0183	5	0.0039	0.0220	13	0.0034	0.0156	-0.0061	0.0128
5	1	427	-0.0077	0.0248	625	-0.0049	0.0257	607	-0.0024	0.0253	-0.0044	-0.0004
5	2	414	-0.0028	0.0217	572	-0.0032	0.0257	588	-0.0043	0.0252	-0.0063	-0.0022
5	3	1317	-0.0092	0.0228	1680	-0.0062	0.0255	1637	-0.0069	0.0249	-0.0081	-0.0057
5	4	3985	-0.0051	0.0229	4533	-0.0036	0.0250	4093	-0.0032	0.0245	-0.0039	-0.0024
5	5	5957	-0.0010	0.0243	4862	0.0020	0.0253	4991	0.0011	0.0246	0.0004	0.0017
5	6	4614	0.0023	0.0246	3087	0.0015	0.0257	3687	0.0011	0.0252	0.0003	0.0019
5	7	2155	0.0038	0.0215	1167	0.0024	0.0234	1633	0.0039	0.0246	0.0027	0.0051
5	8	588	0.0031	0.0197	256	0.0013	0.0194	405	0.0034	0.0223	0.0013	0.0056
5	9	78	0.0022	0.0166	30	0.0021	0.0131	60	0.0042	0.0196	-0.0009	0.0093
5	10	19	0.0054	0.0184	5	0.0120	0.0216	14	0.0099	0.0207	-0.0021	0.0218
6	1	553	-0.0007	0.0237	564	0.0011	0.0255	589	0.0033	0.0292	0.0010	0.0057
6	2	790	0.0002	0.0219	813	0.0003	0.0234	531	0.0011	0.0243	-0.0010	0.0032
6	3	2029	-0.0022	0.0218	2525	-0.0038	0.0231	1507	-0.0026	0.0239	-0.0038	-0.0014
6	4	4410	-0.0006	0.0212	5126	-0.0023	0.0223	3263	-0.0018	0.0231	-0.0026	-0.0010
6	5	5763	0.0004	0.0193	4797	-0.0009	0.0194	4302	-0.0007	0.0219	-0.0013	0.0000
6	6	4138	-0.0002	0.0203	2335	-0.0009	0.0207	3910	0.0010	0.0201	0.0004	0.0017
6	7	1569	0.0010	0.0233	669	-0.0003	0.0234	2447	0.0014	0.0216	0.0005	0.0022
6	8	291	0.0021	0.0232	118	0.0082	0.0252	948	0.0016	0.0243	0.0000	0.0031
6	9	35	0.0073	0.0148	17	0.0000	0.0204	188	0.0038	0.0233	0.0004	0.0071
6	10	6	-0.0086	0.0095	0	0.0000	0.0000	26	0.0049	0.0174	-0.0022	0.0119
7	1	757	0.0013	0.0330	597	-0.0067	0.0281	720	-0.0072	0.0273	-0.0092	-0.0052
7	2	653	0.0055	0.0298	859	-0.0022	0.0279	913	-0.0011	0.0263	-0.0028	0.0006
7	3	1828	-0.0009	0.0309	2342	-0.0023	0.0297	2358	-0.0045	0.0283	-0.0056	-0.0034
7	4	3816	-0.0029	0.0294	4364	-0.0047	0.0293	4326	-0.0049	0.0286	-0.0058	-0.0041
7	5	5196	-0.0034	0.0283	4893	-0.0064	0.0278	4832	-0.0076	0.0271	-0.0083	-0.0068
7	6	4033	-0.0015	0.0271	2896	-0.0035	0.0274	3350	-0.0054	0.0262	-0.0063	-0.0045
7	7	1946	0.0037	0.0248	1108	0.0000	0.0263	1392	0.0005	0.0257	-0.0009	0.0018
7	8	517	0.0090	0.0231	238	0.0053	0.0279	341	0.0089	0.0277	0.0059	0.0118
7	9	91	0.0107	0.0265	45	0.0125	0.0338	43	0.0164	0.0291	0.0074	0.0254
7	10	11	0.0297	0.0206	7	0.0442	0.0381	13	0.0181	0.0186	0.0069	0.0293

8	1	640	0.0040	0.0305	590	-0.0014	0.0268	735	0.0017	0.0274	-0.0003	0.0037
8	2	716	0.0023	0.0267	955	0.0008	0.0245	1077	0.0021	0.0226	0.0007	0.0034
8	3	1740	-0.0018	0.0238	2565	-0.0026	0.0249	2661	-0.0005	0.0240	-0.0014	0.0004
8	4	3652	-0.0043	0.0236	4227	-0.0013	0.0246	4226	-0.0005	0.0246	-0.0012	0.0002
8	5	5147	-0.0057	0.0239	4718	-0.0045	0.0265	4789	-0.0023	0.0261	-0.0030	-0.0016
8	6	4511	-0.0054	0.0244	2994	-0.0028	0.0270	3418	-0.0006	0.0260	-0.0015	0.0003
8	7	2016	-0.0009	0.0279	1030	-0.0004	0.0272	1440	0.0009	0.0265	-0.0004	0.0023
8	8	517	0.0043	0.0262	240	0.0051	0.0224	311	0.0080	0.0233	0.0054	0.0106
8	9	91	0.0050	0.0256	42	0.0095	0.0200	72	0.0088	0.0227	0.0034	0.0141
8	10	22	0.0089	0.0161	7	-0.0064	0.0102	12	0.0135	0.0143	0.0044	0.0226

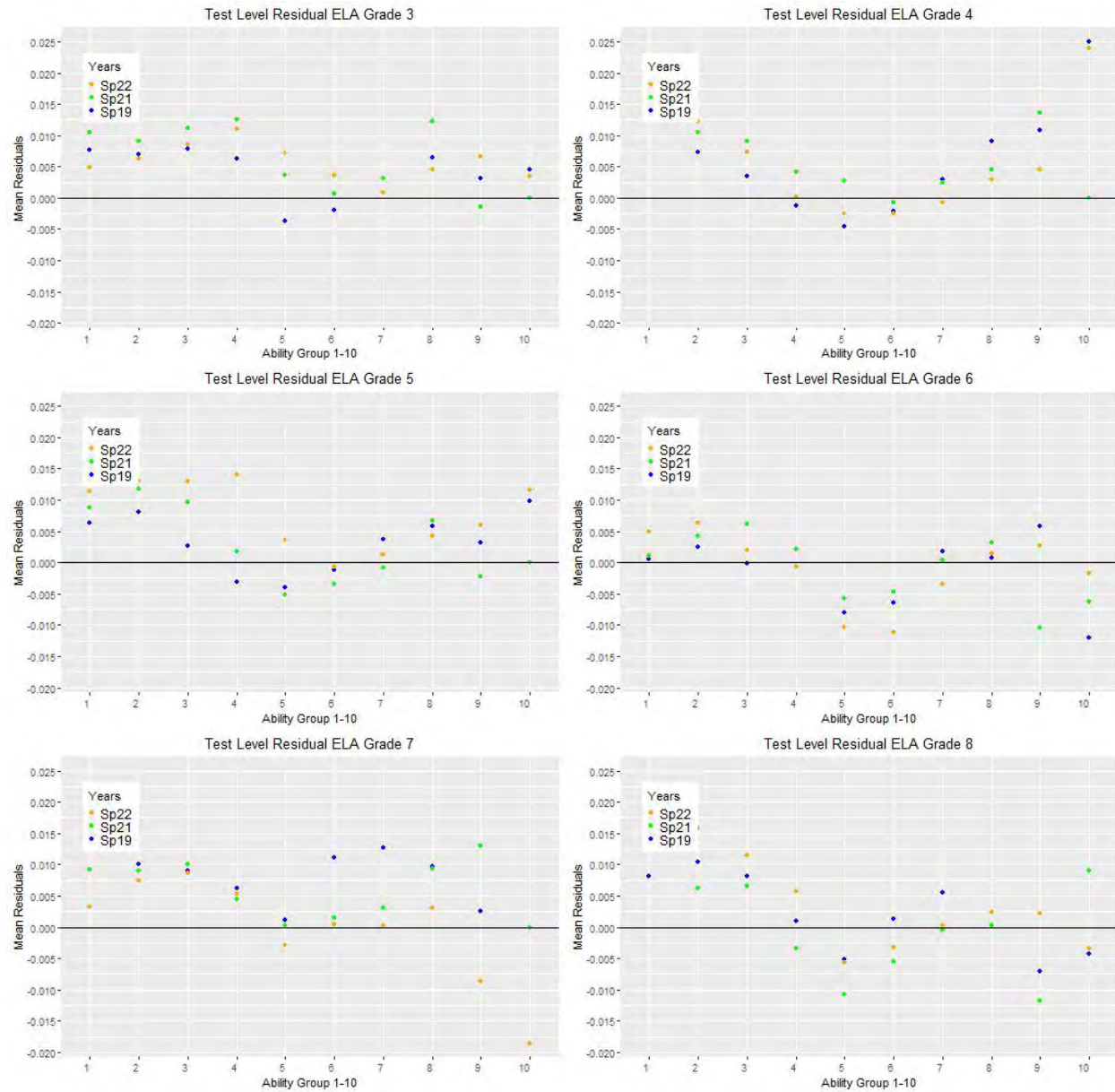


Figure 1: ELA Test Level Residual Comparison

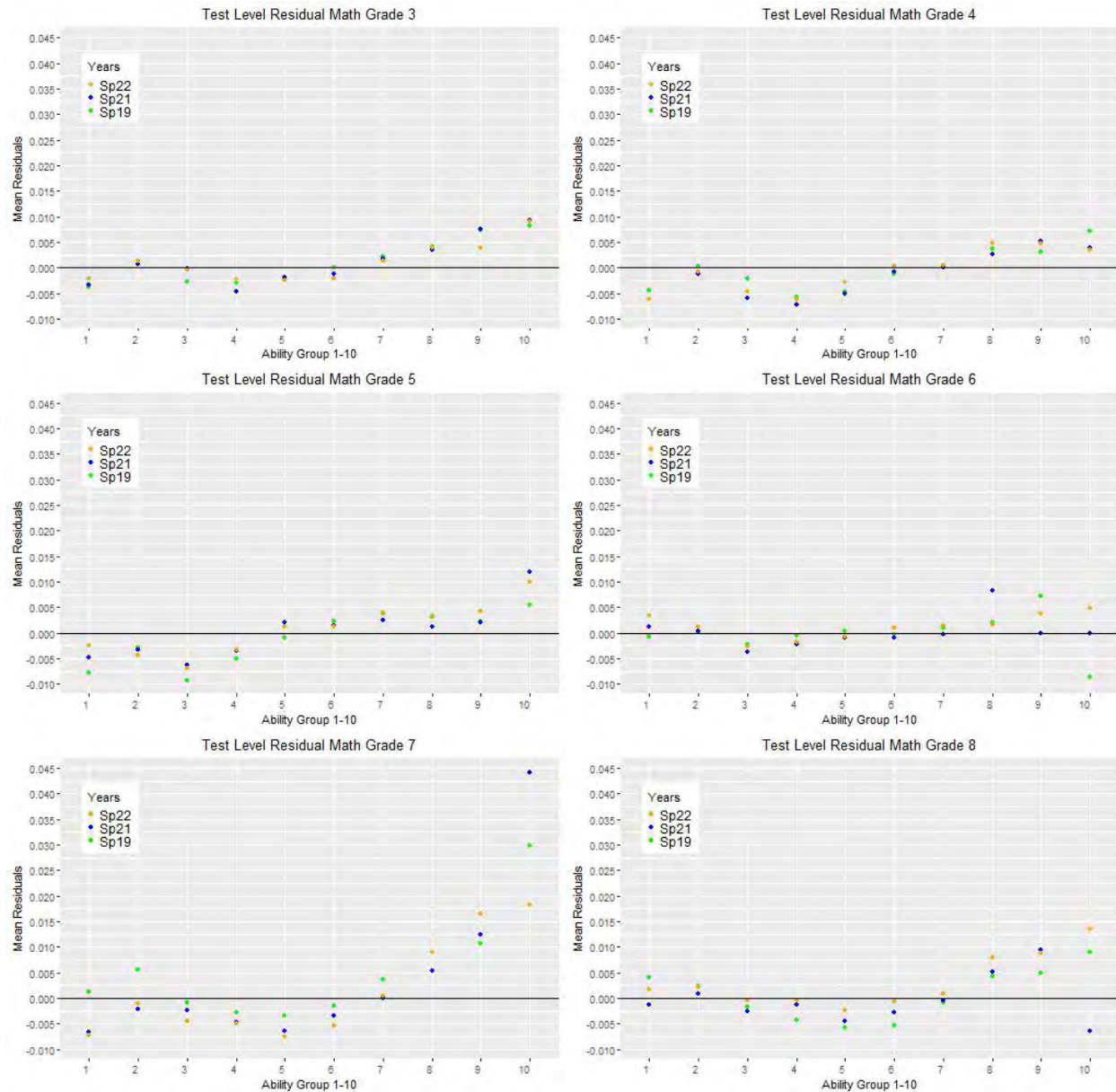


Figure 2: Math Test Level Residual Comparison

6. Correlations

To ensure that the predictive validity of the WVGSA was not impacted by the conditions brought about by the pandemic, correlational studies were conducted between the 2021 and 2022 data both at the individual student level and the aggregate levels, to examine the predictive strength of the scores from 2021 to the scores from 2022. Changes in discriminant validity was also monitored by correlations between ELA and math for 2019, 2021, and 2022. (See Technical Manual, Vol. 4 for a complete discussion). If any of these correlational relationships had been shown to be lower than expected, this would provide potential areas for future research into the extent of the impact of the pandemic on student learning.

Table 15: West Virginia ELA Scale Score Correlations SP18-SP19

SP18_Grade	SP19_Grade	N	Correlation
3	4	18327	0.78
4	5	18693	0.80
5	6	18662	0.80
6	7	17884	0.82
7	8	17984	0.82

Table 16: West Virginia ELA Scale Score Correlations SP19-SP21

SP19_Grade	SP21_Grade	N	Correlation
3	5	15639	0.76
4	6	15637	0.77
5	7	16052	0.79
6	8	16043	0.79

Table 17: West Virginia ELA Scale Score Correlations SP21-SP22

SP21_Grade	SP22_Grade	N	Correlation
3	4	15404	0.78
4	5	15694	0.78
5	6	15624	0.80
6	7	15741	0.81
7	8	16098	0.82

Table 18: West Virginia Math Scale Score Correlations SP18-SP19

SP18_Grade	SP19_Grade	N	Correlation
3	4	18330	0.83
4	5	18699	0.83
5	6	18670	0.81
6	7	17900	0.83
7	8	18007	0.82

Table 19: West Virginia Math Scale Score Correlations SP19-SP21

SP19_Grade	SP21_Grade	N	Correlation
3	5	15698	0.79
4	6	15732	0.79
5	7	16146	0.79
6	8	16147	0.77

Table 20: West Virginia Math Scale Score Correlations SP21-SP22

SP21_Grade	SP22_Grade	N	Correlation
3	4	15408	0.83
4	5	15732	0.83
5	6	15691	0.80
6	7	15855	0.81
7	8	16206	0.80

Table 21: West Virginia Correlations Between ELA and Math Scale Scores

Grade	SP18_N	SP18_Correlation	SP19_N	SP19_Correlation	SP21_N	SP21_Correlation	SP22_N	SP22_Correlation
3	19336	0.75	18545	0.75	16292	0.74	17515	0.75
4	19741	0.74	19199	0.76	16581	0.74	17303	0.75
5	19870	0.74	19533	0.76	16734	0.75	17666	0.76
6	19070	0.78	19564	0.76	16788	0.75	17658	0.76
7	19267	0.75	18812	0.74	17149	0.74	18192	0.75
8	19356	0.74	19013	0.76	17145	0.73	18638	0.75

Results

First, it is important to note that although some differences in performance between the springs of 2021 and 2022 and previous years were detected, variation across testing occasions is expected, and it is difficult to attribute these changes solely to the circumstances brought on by the pandemic. It was not possible to separate all possible sources of variation with the available data. Thus, the following results might best be used to gain insight into areas where additional investigations may be warranted to better understand how and why these changes may have occurred.

Proficiency Levels. Across almost all grades in ELA, the percentage of students that scored above the proficiency cut score was shown to increase since 2021, except for 8th grade students who declined by 4%. Fourth grade students saw the greatest increase with 7% more students being classified as proficient. All grades increased in math proficiency since 2021 ranging from a 3% increase for 7th and 8th grade students to an 8% increase for 3rd grade students.

Scale Scores. Mean scale scores showed similar patterns to those observed in proficiency levels. In ELA, 8th grade students declined in overall scale scores, dropping 5 points from 2021. The remaining grades saw an increase in overall scale scores ranging from 1 point for 7th and 8th grade students to 5 points for 4th grade students. In math, all grades saw an increase in overall scale scores ranging from 3 points for 7th and 8th grade students to 7 points for 4th grade students.

Reliability. Minimal changes in reliability were observed across all grades for both ELA and math. In math, all grades but 3rd and 6th grades showed a slight increase in test reliability ($\alpha = 0.01$), with all other grades maintaining the same reliability coefficient observed in 2021. In ELA, only 4th ($\alpha = 0.02$) and 5th ($\alpha = 0.01$) grade tests showed a slight increase in reliability, with all other grades remaining the same as in 2021.

Item Statistics. Item difficulty differences between the 2021 and 2022 test administration were found to be similar, with far fewer items being flagged than seen in the previous study. In ELA, the percent of items across content standards that demonstrated significant increases in item difficulty since 2021 were found to be low for most standards, ranging from 4% to 50%. The content standards with the greatest number of items flagged since 2021 were found in grades 3 (50%). In math, the percent of content standards with significant increases in item difficulty ranged from 3% to 67%. The math content standards with the greatest number of flagged items were found in grade 8.

Residual Analyses. In general, the residual patterns for spring 2022 follow the patterns found in spring 2019 and spring 2021. In most of the cases, the mean residuals are in the same direction (plus or minus) or are close to zero (the expected average). Considering the variability observed in the past years, the changes between spring 2021 and spring 2022 in cohort groups do not appear to bring concerns related to test score validity/stability of the WVGSA during the pandemic. These non-significant changes in residuals accompanied by the observed declines in test scores (section 1 and section 2) indicate the effectiveness of the CAI CAT item selection algorithm in tailoring item difficulty level to students whose academic performance might have been interrupted by the pandemic.

Correlations. Correlations between the scale scores between spring 2021 and spring 2022 (ranging from $\rho = 0.78$ to $\rho = 0.82$ for ELA, and from $\rho = 0.77$ to $\rho = 0.79$ in math), showed little change compared to the correlations between scale scores between spring 2018 and spring 2019 (ranging from $\rho = 0.78$ to $\rho = 0.82$ for ELA, and from $\rho = 0.80$ to $\rho = 0.83$ in math). Correlations between math and ELA also saw very little change since the previous test administrations.

References

- Boyer, M., Dadey, N., Keng, L. (2020). Statewide summative assessment in spring 2021: A workbook to support planning and decision-making. Dover, NH: National Center for the Improvement of Educational Assessment.
- Council of Chief State School Officers (2020a). Restart and Recovery: Assessments in Spring 2021. Washington, DC: Author.
- Council of Chief State School Officers (2020b). Restart and Recovery: Accountability interrupted: Guidance for collecting, evaluating, and reporting data in 2020-2021. Washington, DC: Author.
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), Educational Measurement (pp. 65 – 110). Westport, CT: Praeger Publishers.