

REPORT

Alignment Analysis of Reading Language Arts Standards and 2008 Pre-Field Test Assessments

**West Virginia
Grades 3-8 and 11
2008**

Norman L. Webb

October 14, 2008

REPORT

Alignment Analysis of Reading Language Arts Standards and 2008 Pre-Field Test Assessments

**West Virginia
Grades 3-8 and 11
2008**

Norman L. Webb

October 14, 2008

Acknowledgements

Reviewers

Grades 3-6

| | | |
|------------------|--------------|-----------|
| Cynthia Jacobson | Group Leader | Wisconsin |
| Bart Wepking | | Wisconsin |
| Joseph Papenfuss | | Wisconsin |
| Dolores LaPoe | | Missouri |
| Pamela Mitchell | | Missouri |
| Cathy Whorton | | Missouri |

Grades 6-8 and 11

| | | |
|------------------------|--------------|-----------|
| John Fortier | Group Leader | Wisconsin |
| Mark Heike | | Wisconsin |
| Elizabeth Vanderschaff | | Minnesota |
| Margaret Bennett | | Missouri |
| Tabatha Beall | | Missouri |
| Sara Sherill | | Missouri |

The State of West Virginia funded this analysis. Timothy Butcher and Beth Judy, Assistant Director, Office Assessment and Accountability, were the main contact persons for the West Virginia Department of Education. Jan Barth, Special Assignment, Superintendent's Office and Director of Office of Assessment and Accountability, had overall responsibility for the study.

Table of Contents

| | |
|---|-----|
| Executive Summary | iii |
| Introduction | 1 |
| Alignment Criteria Used for This Analysis | 3 |
| Categorical Concurrence | 3 |
| Depth-of-Knowledge Consistency | 4 |
| Range-of-Knowledge Correspondence | 7 |
| Balance of Representation | 7 |
| Source of Challenge | 8 |
| Findings | 8 |
| Standards | 8 |
| Alignment of Curriculum Standards and Assessments | 10 |
| Source-of-Challenge Issues and Reviewers' Comments | 23 |
| Reliability Among Reviewers | 23 |
| Summary | 24 |
| References | 26 |
| Appendix A West Virginia Grades 3-8 and 11 Reading Language Arts Standards and Group Consensus DOK Values | |
| Appendix B Data Analysis Tables West Virginia Grades 3-8 and 11 Forms 1 and 2 Reading Language Arts September 2008 | |
| Appendix C Reviewers' Notes and Source-of-Challenge Comments West Virginia Grades 3-8 and 11 Forms 1 and 2 Reading Language Arts September 2008 | |
| Appendix D Debriefing Summary Notes West Virginia Grades 3-8 and 11 Forms 1 and 2 Reading Language Arts September 2008 | |
| Appendix E Data Analysis Tables West Virginia Grades 3-8 and 11 Form TerraNova Reading Language Arts September 2008 | |
| Appendix F Reviewers' Notes and Source-of-Challenge Comments West Virginia Grades 3-8 and 11 | |

Form TerraNova Reading Language Arts September 2008

Appendix G

Debriefing Summary Notes West Virginia Grades 3-8 and 11 Form TerraNova Reading
Language Arts September 2008

Executive Summary

A three-day alignment institute was held September 17-19, 2008 in Charleston, West Virginia to analyze the pre-field test forms of the WESTEST2 and TerraNova with the West Virginia 21st century reading language arts standards. Two groups of six reviewers each participated in the institute. One group analyzed assessments and standards for grades 3-6 and one group analyzed documents for grades 6-8 and 11. Both groups independently analyzed the alignment between the grade 6 Form 1 and attained very similar results. Half of the reviewers were from West Virginia and half were from other states. The reviewers included reading language arts content experts, state reading language arts consultants, reading language arts teachers, and reading language arts assessment experts. Two forms of the WESTEST2 assessment were analyzed for each grade and TerraNova assessments were analyzed for grades 3, 6 and 7.

The alignment for each of the WESTEST2 forms with the West Virginia 21st century reading language arts content standards for grades 3-8 and 11 was mostly acceptable with only minor changes needed to attain full alignment (see summary table below). Four WESTEST2 forms were fully aligned with the 21st century standards—grade 5 Forms 1 and 2 and grade 11 Forms 1 and 2. On these forms, there were only minor issues with balance while all the other criteria were met. The main alignment deficiency for grades 3-5 was that the items were not distributed across enough objectives within the standards along with certain objectives being overemphasized. The criterion for range was not met for grade 3 Forms 1 and 2 and grade 4 Forms 1 and 2. The criterion for balance was not met for Standard I for grade 3 Forms 1 and 2, grade 4 Forms 1 and 2, and for grade 5 Forms 1 and 2. More specifically, in grades 3-4, Objectives RLA O.3.1.08, RLA O.3.2.11, RLA O.4.2.1.08 and RLA O.4.2.11 had a greater proportion of items than other objectives.

For grades 6-8, the main alignment concern was also with the range and balance for writing. The range was not met for Standard II for Forms 1 and 2 for grades 6, 7, and 8. Balance was an issue for writing in grade 8. Otherwise the assessment forms had a sufficient number of items for each standard and DOK consistency. Reviewers found the grade 11 forms to be acceptably aligned. All of the criteria were met on both forms except for a minor balance issue for Standard I on Form 1.

Overall, up to five items on a WESTEST2 pre-field test form would need to be added or replaced to have full alignment. Reviewers did find items in need of more work that were indicated by source-of-challenge comments or in their notes.

The TerraNova forms analyzed for grades 3 and 6 were partially aligned with the West Virginia 21st century reading language arts standards, but the grade 7 TerraNova form needed major improvement. For grade 7 only 13% of the items that targeted objectives under Standard II had a DOK level that was the same or higher than the DOK level of the assigned objective. For grade 6, only slight improvement is needed to attain full alignment. In grade 3, 75% of the writing items were not as complex as the corresponding objectives. Balance was also an issue in this grade, with the items

overemphasizing Objective RLA O.3.1.08. In grade 6, the main issue was with the balance of the writing items, which tended to target Objective RLA O.6.2.06. Reviewers found that the alignment in grade 7 needed improvement. Standard II (Writing) did not meet the criteria for DOK consistency, range and balance. Nearly all of the items were below the DOK of the objectives and one fourth of these items targeted the same objective (RLA O.7.2.04).

Summary Table

Percent of West Virginia Reading Language Arts Standards with Acceptable Level on Each Alignment Criteria for Grade 3-8 and 11 for WESTEST2 Analysis

| Grade | <i>Categorical Concurrence</i> (six or more items) | <i>Depth-of-Knowledge Consistency</i> (50% at/above) | <i>Range of Knowledge</i> (50% of objectives) | <i>Balance of Representation</i> (without possible weakness) | <i>Estimated Range of Items per to be Added or Replaced for Full Alignment</i> |
|----------------|---|---|--|---|--|
| 3 Form 1 | 100 | 100 | 50 | 50 | 1 |
| 3 Form 2 | 100 | 100 | 50 | 50 | 2 |
| 4 Form 1 | 100 | 100 | 50 | 50 | 4 |
| 4 Form 2 | 100 | 100 | 0 | 50 | 5 |
| 5 Form 1 | 100 | 100 | 100 | 0 | 0 |
| 5 Form 2 | 100 | 100 | 100 | 50 | 0 |
| 6 Form 1 | 100 | 100 | 50 | 50 | 1 |
| 6 Form 1 (3-6) | 100 | 100 | 50 | 50 | 1 |
| 6 Form 2 | 100 | 100 | 50 | 50 | 2 |
| 7 Form 1 | 100 | 100 | 50 | 100 | 1 |
| 7 Form 2 | 100 | 100 | 50 | 50 | 1 |
| 8 Form 1 | 100 | 100 | 50 | 50 | 2 |
| 8 Form 2 | 100 | 100 | 50 | 50 | 3 |
| 11 Form 1 | 100 | 100 | 100 | 50 | 0 |
| 11 Form 2 | 100 | 100 | 100 | 100 | 0 |

Categorical Concurrence >6 items
 Depth-of-Knowledge >50% with DOK level the same or higher than level of corresponding Objectives
 Range-of-Knowledge >70% of objectives under a standard
 Balance of Representation A possible weakness if one or more objectives with a relative large number of items (e.g. five or more than the objective with the next highest number of items)

Alignment Analysis of Reading Language Arts Standards and 2008 Pre-Field Test Assessments

West Virginia Grades 3-8 and 11 2008

Norman L. Webb

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 of any one 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).

A three-day Alignment Analysis Institute was conducted September 17-19, 2008, in Charleston, West Virginia. Two groups of six reviewers each analyzed the relationship between items in the pre-field test forms and TerraNova items for the 2008-09 assessments for reading language arts for grades 3-8 and 11. Both groups of reviewers analyzed the grade 6 form 1. Then the two groups were divided, one analyzed the grade 6 Form 2 and one analyzed the grade 6 TerraNova items. After analyzing the grade 6 assessment forms, one group of six reviewers analyzed the grades 3-5 assessments while the other group analyzed the grades 7, 8, and 11 assessments. The language arts reviewers included reading language content experts, state language arts consultants, reading language teachers, and reading language assessment experts.

The State of West Virginia uses the terminology of *content standards* and *objectives* in its reading language arts content expectations. Standards were the broad content requirements across all grades. Three reading language arts standards were included in the analysis for grades 3-8 (reading; writing; and listening, speaking, and media literacy). For grade 11, two reading language arts standards were included in the analysis (reading and writing). Objectives under a standard specified in greater detail what students are to know and do. Each standard had from three to seven objectives. Data for this analysis were entered for the objectives and reported out at the standards level.

As part of the alignment institute, reviewers were trained to identify the depth-of-knowledge of the objectives and assessment items. This training included reviewing the

definitions of the four depth-of-knowledge (DOK) levels and reviewing examples of each. Then the reviewers participated in 1) a review of the depth-of-knowledge levels pre-assigned to the objectives and 2) individual analyses of the assessment items. In reviewing the DOK levels of the objectives, reviewers were instructed not to change any of the DOK values. If the reviewers disagreed with the DOK level assigned to an objective, they were to make a note of this disagreement, but not to change any of the DOK values. The RLA reviewers did not find any DOK levels they disagreed with and did not change any values. Following individual analyses of the items, reviewers participated in a debriefing discussion if time allowed in which they assessed the degree to which they had coded particular items or types of content to the objectives. In addition to the WESTEST2 forms for grades 3-5, all six members of the grades 3-6 group analyzed the grade 3 TerraNova form. All six members of the grades 6-11 group analyzed the grade 7 TerraNova form. Time was not available to analyze the other TerraNova forms (grades 4, 5, 8, and 11).

For each of grades 3 and 11, two WESTEST2 forms for the 2008-2009 assessment were analyzed in this study. In addition, three TerraNova forms were analyzed (grades 3, 6, and 7). The WESTEST2 items were written by WV teachers and CTB staff. Six reviewers participated in the analysis of all forms reviewed.

To derive the results from the analysis, the reviewers' responses were averaged. Any variance among reviewers was considered legitimate, with the true depth-of-knowledge level for the item falling somewhere between two or more assigned values. Such variation could signify a lack of clarity in how the standards and objectives were written, the robustness of an item that can legitimately correspond to more than one objective, and/or a depth of knowledge that falls in between two of the four defined levels. Reviewers were allowed to identify one assessment item as corresponding to up to three objectives—one primary hit (objective) and up to two secondary hits. However, reviewers could only code one depth-of-knowledge level to each assessment item, even if the item corresponded to more than one objective.

Reviewers were instructed to focus primarily on the alignment between the state standards and assessments. However, reviewers were encouraged to offer their opinions on the quality of the standards, or of the assessment activities/items, by writing a note about the item (see Appendices C, D, F, and G). Reviewers could also indicate whether there was a source-of-challenge issue with the item—i.e., a 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.

The results produced from the institute pertain only to the issue of alignment between the West Virginia state standards and the items from the item bank. Note that this alignment analysis does not serve as external verification of the general quality of the state's standards or assessment items. Although reviewers did make a number of suggestions on issues with items or how items could be improved. Only the degree of alignment is discussed in the results. For these results, the means of the reviewers' coding were used to determine whether the alignment criteria were met. When reviewers did

vary in their judgments, the means lessened the error that might result from any one reviewer's finding. Standard deviations are reported in the tables provided in the appendices, which give one indication of the variance among reviewers.

The present report describes the results of an alignment study of objectives and the items for two forms of the pre-field WESTEST2 assessments in grades 3-8 and 11 and three TerraNova forms for grades 3, 6, and 7 in reading language arts for grades 3–8. The study addressed specific criteria related to the content agreement between the state standards and grade-level assessment items. Four criteria received major attention: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation.

Alignment Criteria Used for This Analysis

This analysis judged the alignment between the standards and the assessments on the basis of four criteria. Information is also reported on the quality of items by identifying items with sources-of-challenge and other issues. For each alignment criterion, an acceptable level was defined by what would be required to assure that a student had met the standards.

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 standards and assessment 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 standard. The analysis assumed that one assessment form had to have at least six items for measuring content from a standard in order for an acceptable level of categorical concurrence to exist between the standard and the assessment. The number of items, 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 have to 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 .1, it was estimated that six items would produce an agreement coefficient of at least .63. This indicates that about 63% of the group would be consistently classified as masters or nonmasters if two equivalent test administrations were employed. The agreement coefficient would increase if the cutoff score is increased to one standard deviation from the mean to .77 and, with a cutoff score of 1.5 standard deviations from the mean, to .88. Usually states do not report student results by standards or require students to achieve a specified cutoff score on subscales related to a standard. If a state did do this, then the state would seek a higher agreement coefficient than .63. Six items were assumed as a minimum for an assessment

measuring content knowledge related to a standard, and as a basis for making some decisions about students' knowledge of that standard. If the mean for six items is 3 and one standard deviation is one item, then a cutoff score set at 4 would produce an agreement coefficient of .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. *Depth-of-knowledge consistency between standards and 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 learning goal, as judged in this analysis, at least 50% of the items corresponding to a learning goal had to be at or above the level of knowledge of the learning goal: 50%, a conservative cutoff point, is based on the assumption that a minimal passing score for any one learning goal 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 corresponding learning goal. For example, assume an assessment included six items related to one learning goal 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 standards, 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 learning goal. Some leeway was used in this analysis on this criterion. If a learning goal had between 40% and 50% of items at or above the depth-of-knowledge levels of the standards, then it was reported that the criterion was “weakly” met.

Interpreting and assigning depth-of-knowledge levels to both objectives within standards and assessment items is an essential requirement of alignment analysis. These descriptions help to clarify what the different levels represent in English. Four DOK levels were used to judge both reading and writing standards and the assessment tasks. The reading levels are based on Valencia and Wixson (2000, pp. 909-935). The writing levels were developed by Marshá Horton, Sharon O’Neal, and Phoebe Winter.

Reading

Reading Level 1. Level 1 requires students to receive or recite facts or to use simple skills or abilities. Oral reading that does not include analysis of the text, as well as basic comprehension of a text, is included. Items 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. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Support ideas by 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.

Reading Level 2. Level 2 includes 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 of inference is required. Some important concepts are covered, but not in a complex way. Standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A Level 2 assessment item may require students to apply skills and concepts that are covered in Level 1. However, items require closer understanding of text, possibly through the item's paraphrasing of both the question and the answer. Some examples that represent, but do not constitute all of, Level 2 performance are:

- 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.

Reading Level 3. Deep knowledge becomes a greater focus at Level 3. 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. Standards and items at Level 3 involve reasoning and planning. Students must be able to support their thinking. Items may involve abstract theme identification, inference across an entire passage, or students' application of prior knowledge. Items may also involve more superficial connections between texts. Some examples that represent, but do not constitute all of, Level 3 performance are:

- 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.

Reading Level 4. Higher-order thinking is central and knowledge is deep at Level 4. The standard or assessment item at this level will probably be an extended activity, with extended time provided for completing it. 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. Students take information from at least one passage of a text 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. Some examples that represent, but do not constitute all of, Level 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.

Writing

Writing Level 1. Level 1 requires the student to write or recite simple facts. The focus of this writing or recitation is not on complex synthesis or analysis, but on basic ideas. The students are asked to list ideas or words, as in a brainstorming activity, prior to written composition; are engaged in a simple spelling or vocabulary assessment; or are asked to write simple sentences. Students are expected to write, speak, and edit using the conventions of Standard English. This includes using appropriate grammar, punctuation, capitalization, and spelling. Students demonstrate a basic understanding and appropriate use of such reference materials as a dictionary, thesaurus, or Web site. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Use punctuation marks correctly.
- Identify Standard English grammatical structures, including the correct use of verb tenses.

Writing Level 2. Level 2 requires some mental processing. At this level, students are engaged in first-draft writing or brief extemporaneous speaking for a limited number of purposes and audiences. Students are expected to begin connecting ideas, using a simple organizational structure. For example, students may be engaged in note-taking, outlining, or simple summaries. Text may be limited to one paragraph. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Construct or edit compound or complex sentences, with attention to correct use of phrases and clauses.
- Use simple organizational strategies to structure written work.
- Write summaries that contain the main idea of the reading selection and pertinent details.

Writing Level 3. Level 3 requires some higher-level mental processing. Students are engaged in developing compositions that include multiple paragraphs. These compositions may include complex sentence structure and may demonstrate some synthesis and analysis. Students show awareness of their audience and purpose through focus, organization, and the use of appropriate compositional elements. The use of appropriate compositional elements includes such things as addressing chronological order in a narrative, or including supporting facts and details in an informational report. At this stage, students are engaged in editing and revising to improve the quality of the composition. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Support ideas with details and examples.
- Use voice appropriate to the purpose and audience.
- Edit writing to produce a logical progression of ideas.

Writing Level 4. Higher-level thinking is central to Level 4. The standard at this level is a multi-paragraph composition that demonstrates the ability to synthesize and analyze complex ideas or themes. There is evidence of a deep awareness of purpose and audience. For example, informational papers include hypotheses and supporting evidence. Students are expected to create compositions that demonstrate a distinct voice and that stimulate the reader or listener to consider new perspectives on the addressed ideas and themes. An example that represents, but does not constitute all of, Level 4 performance is:

- Write an analysis of two selections, identifying the common theme and generating a purpose that is appropriate for both.

Range-of-Knowledge Correspondence

For standards 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 standard is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities.* The criterion for correspondence between span of knowledge for a standard and an assessment form considers the number of objectives within the standard with one related assessment item/activity. Fifty percent of the objectives for a standard had to have at least one related assessment item in order 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 standard. This assumes that each benchmark for a standard 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 objective, the requirement that assessment items need to be related to more than 50% of the objectives for a standard increases the likelihood that students will have to demonstrate knowledge on more than one objective per standard 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 objectives. However, any restriction on the number of items included on the test will place an upper limit on the number of objectives that can be assessed. Range-of-knowledge correspondence is more difficult to attain if the content expectations are partitioned among a greater number of standards and a large number of objectives. If 50% or more of the objectives for a standard had a corresponding assessment item, then the range-of-knowledge correspondence criterion was met. If between 40% and 50% of the objectives for a standard had a corresponding assessment item, the criterion was "weakly" met.

Balance of Representation

In addition to comparable depth and breadth of knowledge, aligned standards and assessments require that knowledge be distributed equally in both. The range-of-knowledge criterion only considers the number of objectives within a standard hit (a standard with a corresponding item); it does not take into consideration how the hits (or

assessment items/activities) are distributed among these objectives. *The balance-of-representation criterion is used to indicate the degree to which one objective 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 objectives for a standard that have at least one hit—i.e., one related assessment item per objective. The index is computed by considering the difference in the proportion of objectives and the proportion of hits assigned to the objective. An index value of 1 signifies perfect balance and is obtained if the hits (corresponding items) related to a standard are equally distributed among the objectives for the given standard. Index values that approach 0 signify that a large proportion of the hits are on only one or two of all of the objectives hit. Depending on the number of objectives and the number of hits, a unimodal distribution (most items related to one objective and only one item related to each of the remaining objectives) has an index value of less than .5. A bimodal distribution has an index value of around .55 or .6. Index values of .7 or higher indicate that items/activities are distributed among all of the objectives at least to some degree (e.g., every objective has at least two items) and is used as the acceptable level on this criterion. Index values between .6 and .7 indicate the balance-of-representation criterion has only been “weakly” met.

Source-of-Challenge Criterion

The source-of-challenge criterion is only used to identify items on which the major cognitive demand is inadvertently placed and is other than the targeted reading language arts objective, concept, or application. Cultural bias or specialized knowledge could 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, or at a lower level, even though they possess the understanding and skills being assessed.

Findings

Standards

The DOK levels of the standards assigned prior to this alignment analysis were used. Reviewers reviewed the DOK levels of the objectives and could comment on the level assigned, but did not change any. The DOK levels by grade are summarized in Table 1. The DOK level for each objective is reported in Appendix A. The depth-of-knowledge levels of the objectives progressed in sophistication with the increase in grade. Whereas 37% of the grade 3 objectives were assigned a DOK level 1, less than 5% of the objectives for grade 6 or higher had a DOK level 1. The decline in the objectives with DOK level 1 with an increase in grade was accompanied by an increase in the percentage of objectives assigned a DOK level 3—15% of the grade 3 objectives compared with 50% or more of the objectives for grade 7 or higher. The WV reading language arts objectives clearly increased in complexity over the grades.

Table 1
Percent of Grade-level Expectations for Reading and Writing by Depth-of-Knowledge (DOK) Levels for Grades 3–8 and 11, West Virginia Alignment Analysis for Reading Language Arts 2008 and 2009 Study

| Grade | Total Number of Objectives | DOK Level | Number of objectives by Level | Percent within Standard by Level |
|-------|----------------------------|-----------|-------------------------------|----------------------------------|
| 3 | 27 | 1 | 10 | 37 |
| | | 2 | 11 | 41 |
| | | 3 | 4 | 15 |
| | | 4 | 2 | 7 |
| 4 | 32 | 1 | 7 | 22 |
| | | 2 | 14 | 44 |
| | | 3 | 11 | 34 |
| 5 | 23 | 1 | 4 | 17 |
| | | 2 | 13 | 57 |
| | | 3 | 6 | 26 |
| 6 | 25 | 1 | 1 | 4 |
| | | 2 | 12 | 48 |
| | | 3 | 11 | 44 |
| | | 4 | 1 | 4 |
| 7 | 22 | 2 | 10 | 45 |
| | | 3 | 11 | 50 |
| | | 4 | 1 | 5 |
| 8 | 24 | 1 | 1 | 4 |
| | | 2 | 9 | 38 |
| | | 3 | 14 | 58 |
| 11 | 22 | 1 | 1 | 4 |
| | | 2 | 6 | 27 |
| | | 3 | 12 | 54 |
| | | 4 | 3 | 13 |

If no particular objective is targeted by a given assessment item, reviewers are instructed to code the item at the level of a standard. This coding to a generic objective (standard) sometimes indicates that the item is inappropriate for the grade level. However, if the item is grade-appropriate, then this situation may instead indicate that there is a part of the content not expressly or precisely described in the objectives. These items may highlight areas in the objectives that should be changed, or made more precise. Table 2 displays the assessment items coded to generic objectives by more than one reviewer. Considering all of the items analyzed, two or more reviewers only coded three items from the 14 WESTEST2 forms to generic objectives. This indicates that reviewers were able to find an objective to match nearly all of the items. Reviewers found an additional three items from the TerraNova grade 3 and grade 7 forms that did not map to the West Virginia objectives. For these items, reviewers thought the item targeted content knowledge related to the standard, but not to the specific West Virginia objective.

Reviewers' debriefing comments also highlight some ambiguities in the objectives. These comments can be found in Appendices D and G. Reviewers noted that

Table 2

Items Coded to Generic Objectives by More Than One Reviewer, West Virginia Alignment Analysis for Reading Language Arts, Grades 3-8 and 11 2008

| Grade | Generic Objective | Assessment Item (Number of Reviewers) |
|-----------|-------------------|---------------------------------------|
| 5 Form 1 | RLA.S.1 | 34(5) |
| 5 Form 2 | RLA.S.1 | 36(4) |
| 5 Form 2 | RLA.S.5.1 | 36(2) |
| 3 Form TN | RLA.3.1 | 15(3) |
| 3 Form TN | RLA.S.3.1 | 15(2) |
| 7 Form TN | RLA.S.7.1 | 17(5) |

Alignment of Curriculum Standards and Assessments

Table 3 displays the number of items and points for each assessment form. In the analysis that follows, multiple-point items are given additional weight for alignment purposes. For example, a 3-point item is counted towards the alignment as 3 identically coded 1-point items. All of the items were multiple choice and assigned one point.

Table 3

Number of Items and Point Value by Grade for West Virginia WESTEST2 Pre0Field Test Assessment Forms, Grades 3-8 and 11 2008

| Grade Level | Number of Items | Number of Multi-Point Items | Total Point Value |
|-------------|-----------------|-----------------------------|-------------------|
| 3 Form 1 | 50 | 0 | 50 |
| 3 Form 2 | 50 | 0 | 50 |
| 4 Form 1 | 50 | 0 | 50 |
| 4 Form 2 | 50 | 0 | 50 |
| 5 Form 1 | 50 | 0 | 50 |
| 5 Form 2 | 50 | 0 | 50 |
| 6 Form 1 | 50 | 0 | 50 |
| 6 Form 2 | 50 | 0 | 50 |
| 7 Form 1 | 50 | 0 | 50 |
| 7 Form 2 | 50 | 0 | 50 |
| 8 Form 1 | 50 | 0 | 50 |
| 8 Form 2 | 50 | 0 | 50 |
| 11 Form 1 | 50 | 0 | 50 |
| 11 Form 2 | 50 | 0 | 50 |
| 3 Form TN | 50 | 0 | 50 |
| 6 Form TN | 50 | 0 | 50 |
| 7 Form TN | 60 | 0 | 60 |

The results of the analysis for each of the four alignment criteria are summarized in Tables 4.1-4.18. More detailed data on each of the criteria are given in Appendices B and E in the first three tables. With each table and for each grade, a description of the satisfaction of the alignment criteria for the given grade is provided. The reviewers’ debriefing comments provide further detail about the individual reviewers’ impressions of the alignment.

In Tables 4.1-4.18, “YES” indicates that an acceptable level was attained between the assessment and the learning goal on the criterion. “WEAK” indicates that the criterion was nearly met, within a margin that could simply be due to error in the system. “NO” indicates that the criterion was not met by a noticeable margin—10% over an acceptable level for Depth-of-Knowledge Consistency, 10% over an acceptable level for Range-of-Knowledge Correspondence, and .1 under an index value of .7 for Balance of Representation.

Grade 3

The alignment between the two grade 3 reading language arts assessment forms of the WESTEST2 and the West Virginia 21st century standards for reading and writing was acceptable. At least two items need to be mapped to a standard in order for that standard to be considered assessed. Standard III (Listening, Speaking and Media Literacy) did not have any corresponding items and therefore was considered not to be assessed on both grade 3 forms. Reviewers found the grade 3 Form 1 to be nearly fully aligned except for a balance issue with Standard I (Reading). Sixteen of the thirty-four items mapped to objectives under Standard I all mapped to Objective 3.1.08. The over emphasis of this objective was considered only a minor alignment issue since the other three alignment criteria were acceptable.

Table 4.1
Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 3, Form 1, Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 3, Form 1</i> | <i>Alignment Criteria</i> | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | WEAK |
| RLA.3.2 - WRITING | YES | YES | WEAK | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

NT Less than two items targeted the objective, too few to meaningfully interpret DOK, range, and balance.

The alignment between the grade 3 reading and language arts standards and Form 2 was acceptable. The reviewers found between fifteen and thirty-five items that mapped to Standards I and II. Eighty percent of the items were at or above the DOK level of the

corresponding objectives. Form 2 had a range weakness with Standard II (writing). For this standard, items mapped to only five of the twelve underlying objectives. The balance criterion was not met for Standard I (reading). Similar to Form I, the items overemphasized Objective 3.1.08. The over emphasis of this one objective was considered only a minor alignment issue. Only one for Form 1 and two items for Form 2 would need to be replaced or added to have full alignment. DOK level of the items and the range was appropriate for all standards on both forms.

Table 4.2
Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 3, Form 2, Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 3, Form 1</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | NO |
| RLA.3.2 - WRITING | YES | YES | WEAK | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

Grade 4

The alignment between the grade 4 reading and language arts standards and the two WESTEST2 assessment forms was acceptable. Form 1 of the grade 4 assessment had a sufficient number of items for reading and writing and these items had an appropriate DOK level. The assessment had items that targeted over half of the objectives within Standard I. Standard II had a range weakness due to items targeting only four of the 15 objectives, less than half. Reading had a balance weakness, with 46% of the items targeting one objective. Similar to the previous grade, reviewers found that a large proportion of the items mapped to the objective that asks students to interpret and extend the ideas in literary and informational texts (Objective O.4.1.08). Because the other alignment criteria had acceptable levels for Standard I, the balance weakness is considered more a matter of preference rather than an alignment issue. Standard III (Listening, Speaking and Media Literacy) did not have any corresponding items and therefore was considered not to be assessed on both grade 4 forms.

The alignment between the grade 4 reading language arts standards and Form 2 was similar as for Form 1. The categorical concurrence and DOK criteria were all acceptably met for each standard. The main alignment issue for this form was with the range of objectives targeted by the items. The items targeted less than half of the objectives in both standards. For Standard II, the issue may be related to there being too few items targeting the standard. Reviewers found twelve items total for writing and there are twelve objectives in the standard, making it more difficult to meet the range criteria. Reading had a balance weakness. Again, items that mapped to the reading standard overemphasized objective O.4.1.08.

Overall, to attain full alignment only four items would need to be replaced on Form 1 to target objectives under Standard II that do not have any corresponding items.. For Form 2, five items would need to be added or replaced. Two items on grade 4 Form 1 would need to be replaced to target objectives not currently assessed. Three other items that currently target the overemphasized Objective RLA O.4.1.08 need to be changed to address at least three other objectives under Standard II not currently assessed. Reviewers identified a few items with a source of challenge issue—Form 1 Items 3, 4, 23, and 32 and Form 2 Item 18. Reviewers felt strongly that Form 1 Items 3 and 4 were too ambiguous and had potentially more than one answers.

Table 4.3

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 4, Form 1, Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 4, Form 1</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | NO |
| RLA.3.2 - WRITING | YES | YES | WEAK | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

NT Less that two items targeted the objective, too few to meaningfully interpret DOK, range, and balance.

Table 4.4

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 4 Form 2, Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 4, Form 2</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | WEAK | WEAK |
| RLA.3.2 - WRITING | YES | YES | NO | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

Grade 5

Grade 5 Form 1 and Form 2 were found to be acceptably aligned with the West Virginia reading and writing content standards. Standard III was not assessed. Alignment weaknesses were found on each form for balance. The assessment had an acceptable number of items for each of the reading and writing standards and the DOK levels of the items for both of these standards were at the acceptable level. Range was acceptable for both grade 5 standards on each form, but each form had a weakness in balance. Form 1 overemphasized Objective RLA.O.5.1.08 under Standard I and Objective RLA.O.5.2.05 under Standard II. Form 2 also overemphasized Objective RLA.O.5.1.08. Because the other three alignment criteria were acceptable for each form, a weakness in balance is considered more a preference than an alignment issue for the reading and writing standards. Balance could easily be improved on each form by replacing two or three of the items targeting the objectives that are over emphasized by items that map to other objectives. Even though the grade 5 test forms and standards were found to be acceptably aligned, reviewers found some grade 5 items that had source-of-challenge issues (Form 1 Item 1, 14, 21, 22, 24, 26, 33, and 34 and Form 2 Items 11, 15, 35, and 36). Two of these items did not match any of the objectives, Item 34 in Form 1 and Item 35 in Form 2.

Table 4.5

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 5, Form 1 Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 5, Form 1</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | WEAK |
| RLA.3.2 - WRITING | YES | YES | YES | WEAK |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

NT Less than two items targeted the objective, too few to meaningfully interpret DOK, range, and balance.

Table 4.6

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 5, Form 2 Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 5, Form 2</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | WEAK |
| RLA.3.2 - WRITING | YES | YES | YES | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

Grade 6

The alignment between the grade 6 reading language arts standards and two forms of the WESTEST2 assessments was acceptable with only minor alignment issues. Two groups independently analyzed grade 6 Form 1 (Tables 4.7 and 4.8). The results were essentially the same. The group varied by two items, on the average, coded between the two standards and one writing objective targeted by at least one item. Both groups of reviewers found an issue with the range in Standard II. The difference in balance weakness for Standard I is the result of three reviewers in the grade 6-11 group coding two or three more items to Objective RLA O.6.1.06 than reviewers in the grade 3-6 group. The first group of reviewers judged that the items targeted four, slightly less than half, of the objectives. The second group of reviewers found that the items mapped onto only three of the ten writing objectives and therefore did not meet the range criteria. The balance on Form I was weak for Standard I because of an overemphasis of Objective RLA O.6.1.06. The Categorical Concurrence criterion was acceptable for each standard on both forms along with the DOK Consistency.

Only one item or two items on Form 1 and one item on Form 2 would need to be replaced to attain full alignment between the assessment forms and the grade 6 standards. If one of the items targeting Objective RLA.O.6.1.06 is replaced by an item that targets another objective under Standard I, then this would improve the balance weakness. The range issue can be addressed by adding two items that map to currently untargeted objectives in Standard II. Standard III was not assessed on any of the forms.

Table 4.7

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 6, Form 1 Standards and Assessments for West Virginia Alignment Analysis 2008

| Grade 6, Form 1 | Alignment Criteria | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Standards | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | WEAK |
| RLA.3.2 - WRITING | YES | YES | WEAK | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

NT Less that two items targeted the objective, too few to meaningfully interpret DOK, range, and balance.

Table 4.8

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grades 3-6, Form 1 Standards and Assessments for West Virginia Alignment Analysis 2008

| Grade 6, Form 1 (3-6 Group) | Alignment Criteria | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Standards | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | YES |
| RLA.3.2 - WRITING | YES | YES | NO | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

Table 4.9

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 6, Form 2 Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 6, Form 2</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | WEAK |
| RLA.3.2 - WRITING | YES | YES | WEAK | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

Grade 7

The results from the alignment analysis between the two grade 7 assessment forms and the grade 7 reading language arts standards indicate that the alignment was acceptable. Form 1 of the assessments met the criteria for categorical concurrence, DOK consistency, and balance. Range was fully met for reading but weak for writing. The items that were mapped to the writing objectives targeted slightly less than half of the statements under the standard. Only one reviewer did not coded items to five different objectives under Standard II. On Form 2 the main alignment issue was also with range for Standard II, where items targeted only four of the ten objectives. The balance weakness for Standard I was due to an overemphasis of Objectives RLA.O.7.1.04, RLA O. 7. 1. 05 and RLA O.7.1.07. Because the other alignment criteria had acceptable levels for Standard I, the balance weakness is considered more a matter of preference rather than an alignment issue. Standard III was not assessed.

Overall, one item would need to be replaced or added to each form to have full alignment between the grade 7 West Virginia reading language arts standards and the two WESTEST2 assessment forms. For Form 1 one additional item for Standard II would need to be either added or replaced to target an additional objective for the writing standard. For Form 2 one additional item would be needed for Standard II. These items would need to map onto objectives currently not targeted by the items. Although the alignment was acceptable for both grade 7 forms, reviewers found it difficult to decide which objective an item matched. Often reviewers could not decide whether an item matched Objective RLA O.7.1.07, RLA O.7.1.04 and RLA O.7.1.06.

Table 4.10

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 7, Form 1 Standards and Assessments for West Virginia Alignment Analysis 2008

| Grade 7, Form 1 | Alignment Criteria | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Standards | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | YES |
| RLA.3.2 - WRITING | YES | YES | WEAK | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

NT Less that two items targeted the objective, too few to meaningfully interpret DOK, range, and balance.

Table 4.11

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 7, Form 2 Standards and Assessments for West Virginia Alignment Analysis 2008

| Grade 7, Form 2 | Alignment Criteria | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Standards | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | WEAK |
| RLA.3.2 - WRITING | YES | YES | NO | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

Grade 8

For grade 8 the alignment between both Forms 1 and 2 of the WESTEST2 assessment form and the reading language arts standards was acceptable. The Categorical Concurrence criterion was acceptable for each standard on both forms along with DOK Consistency. The range and balance were also acceptable for Standard I for both forms. On Form 1 range and balance for writing were an issue because of an overemphasis of one objective. Of the 12 items reviewers mapped to Standard II, nine of these items targeted Objective RLA O.8.2.03. Form 2 had similar alignment concerns. Reviewers

found that 78% of the items corresponding to objectives in Standard II mapped to Objective RLA O.8.2.03. For grade 8, balance was an issue because of the over emphasis of the one objective related to a lack of breadth in targeting other objectives. Standard III was not assessed.

Table 4.12

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 8, Form 1 Standards and Assessments for West Virginia Alignment Analysis 2008

| Grade 8, Form 1 | Alignment Criteria | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Standards | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | YES |
| RLA.3.2 - WRITING | YES | YES | NO | WEAK |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

NT Less than two items targeted the objective, too few to meaningfully interpret DOK, range, and balance.

Table 4.13

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 8, Form 2 Standards and Assessments for West Virginia Alignment Analysis 2008

| Grade 8, Form 2 | Alignment Criteria | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Standards | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | YES |
| RLA.3.2 - WRITING | YES | YES | NO | NO |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

For Form 1 to attain full alignment two items would need to be added or replaced to attain full alignment. To address the range and balance issue these items would need to target objectives under Standard II other than Objective RLA O.8.2.03. Similarly, in order for Form 2 to be considered fully aligned, two items would need to be added or replaced so that more objectives are covered under writing. Two or three reviewers noted that two items on the grade Form 1 assessment needed improvement. Reviewers

commented that the answers for Item 32 were not correct and for Item 41 the answers did not correlate directly to the reading passage.

Grade 11

The grade 11 reading language arts standards and two forms of the WESTEST2 were found to be fully aligned. Only one alignment weakness was found with the balance for Standard I on Form 1. Objective O.1.11.06 was overemphasized with sixteen items. Because the other alignment criteria had acceptable levels for Standard I, the balance weakness is considered more a matter of preference rather than an alignment issue. The alignment criteria for all other standards on each of the forms were fully aligned. Only one item that currently maps to Objective O.1.11.06 would need to be replaced by an item that targets another Standard I objective on Form 1 for Form 1 to be fully aligned with the grade 11 standards. Although reviewers judged the grade 11 reading language arts standards and the two assessment forms to be fully aligned, they commented that the overlap between Objectives RLA O.11.1.05, O.11.1.06, O.11.1.08, O.11.1.09 and O.11.1.10 made it difficult to code the items. Reviewers also noted that two items on Form 1 (Item 25 and 38) did not list the correct answer.

Table 4.14

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 11, Form 1, Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 11, Form 1</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | WEAK |
| RLA.3.2 - WRITING | YES | YES | YES | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

NT Less than two items targeted the objective, too few to meaningfully interpret DOK, range, and balance.

Table 4.15

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 11 Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 11, Form 2</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | YES |
| RLA.3.2 - WRITING | YES | YES | YES | YES |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

Grade 3, TerraNova

The alignment between the Terra Nova assessment and the grade 3 standards needed slight improvement. Reviewers judged that 28 items on the assessment targeted objectives under Standard I (Reading) and 26 items under Standard II (Writing). The majority of reviewers did not find any items that targeted objectives under Standard III. At least two items need to be mapped to a standard in order to consider that the objective was assessed. With no items, Standard III was considered not to be assessed by the grade 3 Terra Nova assessment. The DOK Consistency was acceptable for reading. Items targeting Standard 2 were not as complex as expected by the objectives, with 75% of the items below the DOK level of corresponding objectives. The items that were mapped to reading did target a variety of the statements under the standard making the range of items covered by the assessment acceptable. Writing had a range weakness, with less than half of the objectives covered by the items. Balance was an alignment issue for both standards on the grade 3 assessment. For Standard I, 21 out of 28 items mapped to Objective RLA O.3.1.08 and for Standard II, the items overemphasized Objective RLA O.3.2.02.

Overall, the alignment between the grade 3 standards and the TerraNova assessment needed improvement. A total of 8 items would need to be replaced or added to attain full alignment. These items would need to be changed or added to have a similar or higher complexity level as the related objectives to resolve the Depth-of-Knowledge Consistency criterion for Standard II. It is possible for the items that mapped to the Objective RLA O.3.1.08 under Standard I to be replaced by items that targeted more objectives under Standard II in order to address the range and balance issues.

Table 4.16

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade3, Form TN, Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 3, TN</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | NO |
| RLA.3.2 - WRITING | YES | NO | WEAK | WEAK |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

NT Less that two items targeted the objective, too few to meaningfully interpret DOK, range, and balance.

Grade 6, TerraNova

The results from the alignment analysis between the grade 7 Terra Nova assessment and the grade 7 reading language arts standards indicate that the alignment was acceptable. The grade 7 assessment sufficiently covered the reading and writing standards (Standard III was considered not assessed). The assessment items were at an appropriate DOK level for Standard II. Close to half of the items that mapped to Standard I were at or above the same DOK level as the corresponding objectives. The 60 items on the test were distributed across a variety of objectives for each standard making the range acceptable. The balance was good for reading. Standard II had a balance issue because 24 of the 32 items mapped to Objective RLA O. 6.2.06.

To attain full alignment using the minimum acceptable levels only one item would need to be changed. One of the items that currently targets an objective under Standard I needs to be replaced by an item with a higher DOK level. For the alignment to be improved, some of the 24 items that mapped to Objective RLA O.6.2.06 should be redistributed in Standard II to have better balance. Even though the assessment and Standard II are aligned, with 32 items corresponding to Standard II, 75% of the items should not all target one objective. Reviewers identified several items that did not have a corresponding objective (Items 5, 13, 21, and 24). Specifically, reviewers noted that there were no objectives related to persuasive argument, dictionary skills, literal detail and point of view.

Table 4.17

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 6 TN, Grade 3-6 Group, Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 6, TN</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | WEAK | YES | YES |
| RLA.3.2 - WRITING | YES | YES | YES | NO |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

Grade 7, Terra Nova

Alignment between the Grade 7 Terra Nova assessment and the West Virginia content reading language arts standards was found to need major improvement. The Categorical Concurrence, DOK Consistency and Range were met for Standard I. The balance was weakly met but because the other alignment criteria had acceptable levels for Standard I, the balance weakness is considered more a matter of preference rather than an alignment issue. Thus, the grade 7 TerraNova form was aligned with Standard I.

The main alignment issue was with Standard II (Writing). A large proportion of the items mapped to objectives in this standard were not at the same or higher DOK level as the DOK level of the assigned objective. Of the 31 items judged to target objectives under Standard II, 87% of the items were judged to have lower complexity than the complexity of the assigned objective. The range of the items for writing also was not acceptable with items only targeting three of the 10 objectives under Standard II. In addition grade 7 items overemphasized Objective RLA O.7.2.04 causing a balance issue. Of the 31 items that targeted Standard II, 25 items targeted this one objective and all of these items had a DOK level that was lower than the DOK level 3 assigned to this objective.

In order to attain full alignment, a significant number of items would need to be changed. At least twelve items targeting objectives under Standard II would need to be replaced by items that have a higher DOK level. At least two of the new items should target different objectives that do not have any corresponding items. The large number of items that mapped to one objective under Standard II should be more evenly distributed among other objectives to remove the imbalance. These items need to be distributed across at least two more objectives within Standard II, solving the range and balance issue. Reviewers supported these findings by noting that the assessment seemed to heavily test Objective RLA O.7.2.04.

Table 4.18

Summary of Acceptable Levels on Alignment Criteria for Reading Language Arts Grade 7, Form TN, Standards and Assessments for West Virginia Alignment Analysis 2008

| <i>Grade 7, TN</i> | <i>Alignment Criteria</i> | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| <i>Standards</i> | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| RLA.3.1 - READING | YES | YES | YES | WEAK |
| RLA.3.2 - WRITING | YES | NO | NO | NO |
| RLA.3.3 - LISTENING, SPEAKING AND MEDIA LITERACY | NO (N=0) | NT | NT | NT |

Source of Challenge Issue and Reviewers' Comments

Reviewers were instructed to document any source-of-challenge issue and to provide any other comments they may have. These comments can be found in Tables (grade).5 and (grade).7 in Appendices C and F. After coding each grade-level assessment, reviewers also were asked to respond to five debriefing questions. All of the comments made by the reviewers are given in Appendices D and G. The notes in general offer an opinion on the item or give an explanation of the reviewers' coding.

Reliability Among Reviewers

The overall intraclass correlation among the reading language arts reviewers' assignment of DOK levels to items was very high for six reviewers for Grades 3-8 and 11 (Table 5). An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the reviewers. The intraclass correlation was high for each grade, ranging from .82 to .95. A pairwise comparison is used to determine the degree of reliability of reviewers' coding at the objective and standard levels. The standard pairwise comparison values were high and similar to those for other alignment studies. The reviewers did note difficulty in agreeing on which objective to assign certain items. The objective values are high across all grades. Reviewers adjudicated their codings of items to objectives and standards for the WESTEST2 forms, but not for most of the TerraNova forms. The values for the WESTEST2 forms do not represent independent judgments, but the agreement among reviews after some discussion.

Table 5

Intraclass and Pairwise Comparisons, West Virginia Alignment Analysis for Reading Language Arts Grades 3-8 and 11 Assessments

| Grade | Intraclass Correlation | Pairwise Comparison: | Pairwise: Objective | Pairwise: Standard |
|-------------|------------------------|----------------------|---------------------|--------------------|
| 3 Form 1 | .95 | .80 | .84 | .94 |
| 3 Form 2 | .92 | .77 | .85 | .98 |
| 4 Form 1 | .90 | .74 | .90 | .97 |
| 4 Form 2 | .88 | .71 | .90 | .94 |
| 5 Form 1 | .95 | .86 | .90 | .96 |
| 5 Form 2 | .95 | .86 | .82 | .96 |
| 6 Form 1 | .92 | .82 | .67 | .88 |
| 6 Form 2 | .85 | .79 | .64 | .90 |
| 3-6 Form 1 | .92 | .82 | .70 | .92 |
| 7 Form 1 | .82 | .75 | .60 | .90 |
| 7 Form 2 | .83 | .78 | .61 | .91 |
| 8 Form 1 | .90 | .81 | .74 | .92 |
| 8 Form 2 | .86 | .80 | .72 | .96 |
| 11 Form 1 | .90 | .80 | .82 | .94 |
| 11 Form 2 | .88 | .78 | .76 | .97 |
| 3 Form TN | .94 | .82 | .94 | .98 |
| 3-6 Form TN | .90 | .80 | .84 | .96 |
| 7 Form TN | .87 | .79 | .88 | .95 |

Summary

A three-day alignment institute was held September 17-19, 2008 in Charleston, West Virginia to analyze the pre-field test forms of the WESTEST2 and TerraNova with the West Virginia 21st century reading language arts standards. Two groups of six reviewers each participated in the institute. One group analyzed assessments and standards for grades 3-6 and one group analyzed documents for grades 6-8 and 11. Both groups independently analyzed the alignment between the grade 6 Form 1 and attained very similar results. Half of the reviewers were from West Virginia and half were from other states. The reviewers included reading language arts content experts, state reading language arts consultants, reading language arts teachers, and reading language arts assessment experts. Two forms of the WESTEST2 assessment were analyzed for each grade and TerraNova assessments were analyzed for grades 3, 6 and 7.

The alignment for each of the WESTEST2 forms with the West Virginia 21st century reading language arts content standards for grades 3-8 and 11 was mostly acceptable with only minor changes needed to attain full alignment (see summary table below). Four WESTEST2 forms were fully aligned with the 21st century standards—grade 5 Forms 1 and 2 and grade 11 Forms 1 and 2. On these forms, there were only

minor issues with balance while all the other criteria were met. The main alignment deficiency for grades 3-5 was that the items were not distributed across enough objectives within the standards along with certain objectives being overemphasized. The criterion for range was not met for grade 3 Forms 1 and 2 and grade 4 Forms 1 and 2. The criterion for balance was not met for Standard I for grade 3 Forms 1 and 2, grade 4 Forms 1 and 2, and for grade 5 Forms 1 and 2. More specifically, in grades 3-4, Objectives RLA O.3.1.08, RLA O.3.2.11, RLA O.4.2.1.08 and RLA O.4.2.11 had a greater proportion of items than other objectives.

For grades 6-8, the main alignment concern was also with the range and balance for writing. The range was not met for Standard II for Forms 1 and 2 for grades 6, 7, and 8. Balance was an issue for writing in grade 8. Otherwise the assessment forms had a sufficient number of items for each standard and DOK consistency. Reviewers found the grade 11 forms to be acceptably aligned. All of the criteria were met on both forms except for a minor balance issue for Standard I on Form 1.

Overall, up to five items on a WESTEST2 pre-field test form would need to be added or replaced to have full alignment. Reviewers did find items in need of more work that were indicated by source-of-challenge comments or in their notes.

The TerraNova forms analyzed for grades 3 and 6 were partially aligned with the West Virginia 21st century reading language arts standards, but the grade 7 TerraNova form needed major improvement. For grade 7 only 13% of the items that targeted objectives under Standard II had a DOK level that was the same or higher than the DOK level of the assigned objective. For grade 6, only slight improvement is needed to attain full alignment. In grade 3, 75% of the writing items were not as complex as the corresponding objectives. Balance was also an issue in this grade, with the items overemphasizing Objective RLA O.3.1.08. In grade 6, the main issue was with the balance of the writing items, which tended to target Objective RLA O.6.2.06. Reviewers found that the alignment in grade 7 needed improvement. Standard II (Writing) did not meet the criteria for DOK consistency, range and balance. Nearly all of the items were below the DOK of the objectives and one fourth of these items targeted the same objective (RLA O.7.2.04).

Summary Table

Percent of West Virginia Reading Language Arts Standards with Acceptable Level on Each Alignment Criteria for Grade 3-8 and 11 for WESTEST2 Analysis

| Grade | <i>Categorical Concurrence</i> (six or more items) | <i>Depth-of-Knowledge Consistency</i> (50% at/above) | <i>Range of Knowledge</i> (50% of objectives) | <i>Balance of Representation</i> (without possible weakness) | <i>Estimated Range of Items per to be Added or Replaced for Full Alignment</i> |
|----------------|---|---|--|---|--|
| 3 Form 1 | 100 | 100 | 50 | 50 | 1 |
| 3 Form 2 | 100 | 100 | 50 | 50 | 2 |
| 4 Form 1 | 100 | 100 | 50 | 50 | 4 |
| 4 Form 2 | 100 | 100 | 0 | 50 | 5 |
| 5 Form 1 | 100 | 100 | 100 | 0 | 0 |
| 5 Form 2 | 100 | 100 | 100 | 50 | 0 |
| 6 Form 1 | 100 | 100 | 50 | 50 | 1 |
| 6 Form 1 (3-6) | 100 | 100 | 50 | 50 | 1 |
| 6 Form 2 | 100 | 100 | 50 | 50 | 2 |
| 7 Form 1 | 100 | 100 | 50 | 100 | 1 |
| 7 Form 2 | 100 | 100 | 50 | 50 | 1 |
| 8 Form 1 | 100 | 100 | 50 | 50 | 2 |
| 8 Form 2 | 100 | 100 | 50 | 50 | 3 |
| 11 Form 1 | 100 | 100 | 100 | 50 | 0 |
| 11 Form 2 | 100 | 100 | 100 | 100 | 0 |

| | |
|---------------------------|--|
| Categorical Concurrence | >6 items |
| Depth-of-Knowledge | >50% with DOK level the same or higher than level of corresponding Objectives |
| Range-of-Knowledge | >70% of objectives under a standard |
| Balance of Representation | A possible weakness if one or more objectives with a relative large number of items (e.g. five or more than the objective with the next highest number of items) |

References

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.

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.

