

REPORT

Alignment Analysis of Mathematics Standards and 2008 Pre-Field Test Assessments

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

Norman L. Webb

October 5, 2008

REPORT

Alignment Analysis of Mathematics Standards and 2008 Pre-Field Test Assessments

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

Norman L. Webb

October 5, 2008

Acknowledgements

Reviewers:

Group Grades 6-8 and 11

Linda McQuillen	(Group Leader)	WI
Bill Hadley		PA
Linda Wilson		DC
Natalie Zimmerman		WV
Judy Pomeroy		WV
Sue Steinbeck		WV

Group Grades 3-6

Lynn Raith	(Group Leader)	PA
Mary Derginer		WI
Jennifer Theisen		WI
Brian Perkins		WV
Mary Jane Merendino		WV
Holly Woods		WV

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	iv
Introduction.....	1
Alignment Criteria Used for This Analysis	3
Categorical Concurrence.....	3
Depth-of-Knowledge Consistency.....	4
Range-of-Knowledge Correspondence.....	6
Balance of Representation	6
Source of Challenge.....	7
Findings.....	7
Standards.....	7
Alignment of Curriculum Standards and Assessments.....	9
Source-of-Challenge Issues and Reviewers’ Comments	24
Reliability Among Reviewers.....	24
Summary	26
References.....	28
Appendix A	
West Virginia Grades 3-8 and 11 Mathematics Standards and Group Consensus DOK Values	
Appendix B	
Data Analysis Tables West Virginia Grades 3-8 and 11 WESTEST2 Forms 1 and 2 Mathematics 2008	
Appendix C	
Reviewers’ Notes and Source-of-Challenge Comments West Virginia Grades 3-8 and 11 WESTEST2 Forms 1 and 2 Mathematics 2008	
Appendix D	
Debriefing Summary Notes West Virginia Grades 3-8 and 11 WESTEST2 Forms 1 and 2 Mathematics 2008	
Appendix E	
Data Analysis Tables West Virginia Grades 3-8 and 11 TerraNova Forms Mathematics 2008	

Appendix F
Reviewers' Notes and Source-of-Challenge Comments West Virginia Grades 3-8 and 11
TerraNova Forms Mathematics 2008

Appendix G
Debriefing Summary Notes West Virginia Grades 3-8 and 11 TerraNova Forms
Mathematics 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 mathematics 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 these documents for grades 6-8 and 11. Both groups independently analyzed the alignment with grade 6 Form 1 and attained the same results. Half of the reviewers were from West Virginia and half were from other states. The reviewers included mathematics education content experts, district mathematics supervisors, mathematics teachers, and mathematics assessment experts. Two forms of the WESTEST2 assessment and one TerraNova form were analyzed for each grade.

The alignment for each of the WESTEST2 forms with the West Virginia 21st century mathematics content standards for grades 3-8 and 11 was at least acceptable with only minor changes needed to attain full alignment (see summary table below). Three WESTEST2 forms were fully aligned with the 21st century standards—grade 3 Form 1, grade 6 Form 1, and grade 7 Form 2. The main alignment deficiency for grades 3-5 was fewer than six items, the acceptable level for the Categorical Concurrence criterion, for either or both Standards II (Algebra) and V (Data Analysis and Probability). The depth-of-knowledge (DOK) levels of the items, except for grade 5 Form 2, were suitable along with the range. One or two objectives had a greater proportion of items than other objectives for one standard for grade 3 Forms 1 and 2 and grade 4 Form 1. These balance weaknesses were considered more a matter of preference since the other alignment criteria for the standard were acceptable.

For grades 6-8, the main alignment concern was with the DOK levels of items for one or two standards—Standard I (Number and Operations) for grade 6 Form 2 and two grade 8 forms, Standard II (Algebra) grade 7 Form 1, and Standard IV (Measurement) grade 6 Form 2. Otherwise the assessment forms had a sufficient number of items for each standard (except Standard IV grade 8 Form 2), range, and balance.

Reviewers found an insufficient number of items on both of the grade 11 forms for the trigonometry standard, only two or three items. These items also were lower in complexity than the assigned objectives. Range and balance both were suitable.

Overall, up to four 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, particularly grade 11 Form 1.

The TerraNova forms for all seven grades analyzed were not aligned with the West Virginia 21st century mathematics standards. The TerraNova forms for grades 3-8 had at least some items for each of the five mathematics standards. From one-third (grade 7) to two-thirds (grade 5) of the items on each form mapped to Standard I (Number and

Operations). The other four standards generally had less than six items for each of the grades. The DOK levels of the items were low for two to four standards for each grade. Range was poor for two to four of the standards for each grade. Balance for Standard I for grades 5, 6, and 8 was weak because one objective was over emphasized.

Reviewers did not find any items on the grade 11 TerraNova form that targeted the Algebra II or trigonometry standards. The DOK levels of the items that targeted the other two standards were low. Range for the Algebra I standard was barely acceptable and not acceptable for the geometry standard.

Summary Table

Percent of West Virginia Mathematics 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	100	80	0
3 Form 2	60	100	100	80	2
4 Form 1	60	100	100	80	2
4 Form 2	80	100	100	100	1
5 Form 1	60	100	100	100	3
5 Form 2	80	80	100	100	3
6 Form 1	100	100	100	100	0
6 Form 2	100	60	100	80	3
7 Form 1	100	80	100	100	1
7 Form 2	100	100	100	100	0
8 Form 1	100	80	100	100	1
8 Form 2	80	80	100	100	2-3
11 Form 1	75	75	100	100	4
11 Form 2	75	75	100	100	3

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 Mathematics Standards and 2008 Pre-Field Test Assessments

West Virginia Grades 3-8 and 11 2008

Norman L. Webb

Introduction

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 mathematics 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 mathematics reviewers included mathematics content experts, district mathematics supervisors, mathematics teachers, and mathematics assessment experts.

The State of West Virginia uses the terminology of *content standards* and *objectives* in its mathematics content expectations. Standards were the broad content requirements across all grades. Five mathematics standards were included in the analysis for grades 3-8 (number and operations, algebra, geometry, measurement, and data analysis and probability). For grade 11, four mathematics standards were included in the analysis (algebra 1, algebra 2, geometry, and trigonometry). Objectives specified in greater detail under a standard what students are to know and do. Each standard had from two to 17 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 mathematics 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. Because of the large volume of work, for some assessment parts (grades 3-6 and 8), two reviewers conducted the analysis.

Items for the 2008-2009 assessment organized into three forms were analyzed in this study. The items written by WV teachers and CTB staff and the TerraNova items for a grade level were included in the analysis. Two forms were those items written by West Virginia teachers and CTB staff culled from the item bank. The third form for each grade was composed of TerraNova items. Forms 1 and 2 were analyzed by all of the reviewers. The TerraNova form for each grade was analyzed by two or three reviewers.

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 tests and TerraNova items by at least two reviewers for each grade in mathematics for grades 3–8 and 11 in West Virginia. 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 an assessment form and the standard, as judged in this analysis, at least 50% of the items corresponding to a standard had to be at or above the level of knowledge of the standard: 50%, a conservative cutoff point, is based on the assumption that a minimal passing score for any one standard 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 standard. For example, assume an assessment included six items related to one standard 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 standard. Some leeway was used in this analysis on this criterion. If a standard 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 to assessment items is an essential requirement of alignment analysis. These descriptions help to clarify what the different levels represent in mathematics:

Level 1 (Recall) includes the recall of information such as a fact, definition, term, or a simple procedure, as well as performing a simple algorithm or applying a formula. That is, in mathematics, a one-step, well defined, and straight algorithmic procedure should be included at this lowest level. Other key words that signify Level 1 include “identify,” “recall,” “recognize,” “use,” and “measure.” Verbs such as “describe” and “explain” could be classified at different levels, depending on what is to be described and explained.

Level 2 (Skill/Concept) includes the engagement of some mental processing beyond an habitual response. A Level 2 assessment item requires students to make some decisions as to how to approach the problem or activity, whereas Level 1 requires students to demonstrate a rote response, perform a well-known algorithm, follow a set procedure (like a recipe), or perform a clearly defined series of steps. Keywords that generally distinguish a Level 2 item include “classify,” “organize,” “estimate,” “make observations,” “collect and display data,” and “compare data.” These actions imply more than one step. For example, to compare data requires first identifying characteristics of the objects or phenomenon and then grouping or ordering the objects. Some action verbs, such as “explain,” “describe,” or “interpret,” could be classified at different levels depending on the object of the action. For example, interpreting information from a simple graph, or requiring mathematics information from the graph, also is at Level 2. Interpreting information from a complex graph that requires some decisions on what features of the graph need to be considered and how information from the graph can be aggregated is at Level 3. Level 2 activities are not limited solely to number skills, but can involve visualization skills and probability skills. Other Level 2 activities include noticing and describing non-trivial patterns; explaining the purpose and use of experimental procedures; carrying out experimental procedures; making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts.

Level 3 (Strategic Thinking) requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. In most instances, requiring students to explain their thinking is at Level 3. Activities that require students to make conjectures are also at this level. The cognitive demands at Level 3 are complex and abstract. The complexity does not result from the fact that there are multiple answers, a possibility for both Levels 1 and 2, but because the task requires more demanding reasoning. An activity, however, that has more than one possible answer and requires students to justify the response they give would most likely be at Level 3. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve problems.

Level 4 (Extended Thinking) requires complex reasoning, planning, developing, and thinking, most likely over an extended period of time. 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, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as Level 2. However, if the student is to conduct a river study that requires taking into consideration a number of variables, this would be at Level 4. At Level 4, the cognitive demands of the task should be high and the work should be very complex. Students should be required to make several connections—relate ideas *within* the content area or *among* content areas—and to select one approach among many alternatives on how the situation should be solved, in order to be at this highest level. Level 4 activities include developing and proving conjectures; designing and conducting experiments; making connections between a finding and

related concepts and phenomena; combining and synthesizing ideas into new concepts; and critiquing experimental designs.

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 mathematics 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 majority of the DOK levels of the WV Content Standards and Objectives (CSOs) was a DOK level 2 (skills and concepts). However, each grade from grade 3 to 8 had at least one objective at each DOK level indicating that students were expected to do analysis and extended thinking performance at all grades. The level of complexity increased in complexity and sophistication across the grades. The proportion of objectives with a DOK level 3 or 4 increased to over one-third of the objectives for grade 8. The majority of grade 11 objectives was a DOK level 2, as for the lower grades, but did not have an objective judged to have a DOK level 4. About one-fourth of the grade 11 objectives were judged to have a DOK level 3.

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” 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 six 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 more items from the TerraNova that did not map to the West Virginia objectives. Two or more reviewers coded 24 TerraNova items to generic 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 E, H, K, and N.

Table 1

Percent of Grade-level Expectations by Depth-of-Knowledge (DOK) Levels for Grades 3–8 and 11, West Virginia Alignment Analysis for Mathematics 2005 and 2006 Study

Grade	Total Number of Objectives	DOK Level	Number of objectives by Level	Percent within Standard by Level
3	34	1	12	35
		2	17	50
		3	4	11
		4	1	2
4	29	1	11	37
		2	14	48
		3	3	10
		4	1	3
5	32	1	8	25
		2	21	65
		3	2	6
		4	1	3
6	31	1	6	19
		2	20	64
		3	4	12
		4	1	3
7	29	1	6	20
		2	13	44
		3	9	31
		4	1	3
8	27	1	4	14
		2	13	48
		3	9	33
		4	1	3
11	38	1	5	13
		2	23	60
		3	10	26

Table 2

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

Grade	Generic Objective	Assessment Item (Number of Reviewers)
3 Form 2	M.S.3.2	22 (5)
3 Form 2	M.S.3.4	20 (6)
6 Form 1	M.S.6.4	17 (6)
6 Form 2	M.S.6.4	17(6)
7 Form 1	M.S.7.2	21 (2)
7 Form 2	M.S.7.2	20 (6)
4 Form TN	M.S.4.1	1 (2)
5 Form TN	M.S.5.1	3 (2); 6 (2); 21 (2)
5 Form TN	M.S.5.2	28 (2)
5 Form TN	M.S.5.3	17 (2)
6 Form TN	M.S.6.1	2 (2); 13 (3); 15 (2)
6 Form TN	M.S.6.1	2 (2); 13 (3); 15 (4); 29 (4)
6 Form TN	M.S.6.2	1 (2)
6 Form TN	M.S.6.3	28 (4); 30 (6)
6 Form TN	M.S.6.4	22 (2)
7 Form TN	M.S.7.3	31 (3)
7 Form TN	M.S.7.5	10 (2); 11 (2)
8 Form TN	M.S.8.1	26 (2); 28 (2)
8 Form TN	M.S.8.2	27 (2)
8 Form TN	M.S.8.4	5 (2)

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 for all grades were assigned a point value of one. The total points assigned to each form equaled the total number of items on each form.

The results of the analysis for each of the four alignment criteria are summarized in Tables 4.1-11.7. 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 through 11.7, "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.

Table 3

Number of Items and Point Value by Grade for West Virginia Assessments, Grades 3-8 and 11 2008

Grade Level	Number of Items	Number of Multi-Point Items	Total Point Value
3 Form 1	45	1	45
3 Form 2	45	1	45
4 Form 1	45	1	45
4 Form 2	45	1	45
5 Form 1	45	1	45
5 Form 2	45	1	45
6 Form 1	45	1	45
6 Form 2	45	1	45
3-6 Form 1	45	1	45
7 Form 1	45	1	45
7 Form 2	45	1	45
8 Form 1	45	1	45
8 Form 2	45	1	45
11 Form 1	45	1	45
11 Form 2	45	1	45
3 Form TN	30	1	30
4 Form TN	30	1	30
5 Form TN	30	1	30
6 Form TN	30	1	30
7 Form TN	32	1	32
8 Form TN	31	1	31
11 Form TN	25	1	25

Grade 3

The alignment between the two grade 3 mathematics assessment forms of the WESTEST2 and the West Virginia standards was acceptable. Form 1 and the standards were considered fully aligned while the alignment with Form 2 was acceptable. The grade 3 form was found to be nearly fully aligned except for a balance issue with Standard II. Four of the six items mapped to objectives under Standard II all mapped to Objective 3.2.1. The over emphasis of this objective was considered only a minor alignment issue.

The alignment between the grade 3 mathematics standards and Form 2 was acceptable. Not all reviewers agreed that one of the six items that mapped to Standard II and V corresponded to that standard. As a result fewer than six items were judged to be mapped to objectives under these two standards. Form 2 also had a balance weakness with Standard IV. For this standard, four of the six or seven items that mapped to underlying objectives were judged to target the same objective, Objective 3.4.2. The over emphasis of this one objective was considered only a minor alignment issue. The DOK level of the items and the range was appropriate for all standards on both forms.

Table 4.1

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 3, Form 1, Standards and Assessments for West Virginia Alignment Analysis 2008

Grade 3, 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>
M.S.3.I - NUMBER AND OPERATIONS	YES	YES	YES	YES
M.S.3.II - ALGEBRA	YES	YES	YES	WEAK
M.S.3.III - GEOMETRY	YES	YES	YES	YES
M.S.3.IV - MEASUREMENT	YES	YES	YES	YES
M.S.3.V - DATA ANALYSIS AND PROBABILITY	YES	YES	YES	YES

Table 4.2

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 3, Form 2, Standards and Assessments for West Virginia Alignment Analysis 2008

Grade 3 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>
M.S.3.I - NUMBER AND OPERATIONS	YES	YES	YES	YES
M.S.3.II - ALGEBRA	NO (5.5)	YES	YES	YES
M.S.3.III - GEOMETRY	YES	YES	YES	YES
M.S.3.IV - MEASUREMENT	YES	YES	YES	WEAK
M.S.3.V - DATA ANALYSIS AND PROBABILITY	NO (5.83)	YES	YES	YES

Overall, the reviewers agreed that the alignment for grade 3 was acceptable. Two additional items would be needed to have full alignment for grade 3 Form 2—one item for Standard II and one item for Standard V. The balance issues were considered a matter of preference, issues that could be resolved by replacing one or two items. The overemphasis of one objective was not considered too critical because all of the other alignment criteria were acceptably met. Two or more reviewers did identify two items on

Form 1 (Items 18 and 26) and one item on Form 2 (Item 28) with source-of-challenge issues. Three reviewers reported that some topics were not addressed including different models for subtraction. In general, the alignment was good for grade 3.

Grade 4

The alignment between the grade 4 mathematics standards and the two WESTEST2 assessment forms was acceptable. Only one or two items would need to be replaced or added to have full alignment. Form 1 of the grade 4 assessment did not meet the acceptable level of six items per standard for the Categorical Concurrence criterion. One of the six reviewers did not agree that six items targeted either Standard IV or Standard V resulting in an average number of items slightly fewer than six items. Otherwise, the items on Form 1 had an appropriate DOK level, range, and balance. The balance weakness for Standard I is because reviewers found six items that mapped to Objective M.O.4.1.7 and four items that mapped to Objective M.O.4.1.7 with only none to two items mapped to the other eight objectives. 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 between the grade 4 mathematics standards and Form 2 was similar as for Form 1. The DOK, range, and balance criteria were all acceptably met for each of the five standards. The only alignment issue was with Standard IV that only had, on the average, 5.83 items that targeted underlying objectives.

The alignment between the grade 4 two assessment forms and the standards was acceptable. Only one or two items would need to be replaced to have full alignment. Reviewers identified a few items with a source of challenge issue—Form 1 Items 22, 24, and 25 and Form 2 Item 13. Reviewers felt strongly that Form 1 Items 24 and 25 were too difficult.

Table 5.1
Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 4, Form 1, Standards and Assessments for West Virginia Alignment Analysis 2008

Grade 4 Form 1 <i>Standards</i>	Alignment Criteria			
	<i>Categorical Concurrence</i>	<i>Depth-of-Knowledge Consistency</i>	<i>Range of Knowledge</i>	<i>Balance of Representation</i>
M.S.4.I - NUMBER AND OPERATIONS	YES	YES	YES	WEAK
M.S.4.II - ALGEBRA	YES	YES	YES	YES
M.S.4.III - GEOMETRY	YES	YES	YES	YES
M.S.4.IV - MEASUREMENT	NO (5.83)	YES	YES	YES
M.S.4.V - DATA ANALYSIS AND PROBABILITY	NO (5.83)	YES	YES	YES

Table 5.2

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

<i>Grade 4 Form 2</i> <i>Standards</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>
M.S.4.I - NUMBER AND OPERATIONS	YES	YES	YES	YES
M.S.4.II - ALGEBRA	YES	YES	YES	YES
M.S.4.III - GEOMETRY	YES	YES	YES	YES
M.S.4.IV - MEASUREMENT	NO (5.83)	YES	YES	YES
M.S.4.V - DATA ANALYSIS AND PROBABILITY	YES	YES	YES	YES

Grade 5

The alignment between the grade 5 mathematics standards and two forms of the WESTEST2 assessments was acceptable with only minor alignment issues. Form 1 of the grade 5 assessment had an acceptable number of items for three of the five standards. For Standard II (Algebra) reviewers only found five items and for Standard V (Data Analysis and Probability) all reviewers could only agree on four items. The DOK level, range, and balance were all good for Form 1. Form 2 had a sufficient number of items for four of the five standards. Reviewers only found three items that targeted content knowledge under Standard V. Of these three items that mapped to Standard V, only one item had an appropriate DOK level. The Depth-of-Knowledge Consistency criterion was acceptably met for the four standards, but not for Standard V. Range and balance was acceptable for all five standards.

Overall, three items would need to be replaced or added to both forms to have full alignment between the grade 5 West Virginia mathematics standards and the two WESTEST2 assessment forms. For Form 1 one additional item for Standard II and two additional items for Standard V would be needed to have at least six items for each of these standards. For Form 2 three additional items would be needed for Standard V. All of these additional items would need to have an appropriate DOK level. One reviewer identified some content topics not included on Form 1:

Number and Operations: Model and write equivalencies of fractions, decimals, percents, and ratios.

Algebra: solve simple equations and inequalities using patterns and models of real-world situations, create graphs on number lines of the equations and interpret the results.

Measurement: develop strategies to determine the volume of a rectangular prism.

Another reviewer noted topics not addressed by Form 2, “Did not see many items that addressed modeling and writing equivalent/nonequivalent fractions, decimals, percents, and ratios.” One reviewer cautioned about Form 1, “The wording on items should be

looked at to determine what is really being assessed and how much information (extraneous) needs to be in a problem to make it real life and how much of it will only serve to confuse the students.”

Table 6.1

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

Grade 5 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>
M.S.5.I - NUMBER AND OPERATIONS	YES	YES	YES	YES
M.S.5.II - ALGEBRA	NO (5.17)	YES	YES	YES
M.S.5.III - GEOMETRY	YES	YES	YES	YES
M.S.5.IV - MEASUREMENT	YES	YES	YES	YES
M.S.5.V - DATA ANALYSIS AND PROBABILITY	NO (4.83)	YES	YES	YES

Table 6.2

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

Grade 5, 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>
M.S.5.I - NUMBER AND OPERATIONS	YES	YES	YES	YES
M.S.5.II - ALGEBRA	YES	YES	YES	YES
M.S.5.III - GEOMETRY	YES	YES	YES	YES
M.S.5.IV - MEASUREMENT	YES	YES	YES	YES
M.S.5.V - DATA ANALYSIS AND PROBABILITY	NO (3.5)	NO	YES	YES

Grade 6

Grade 6 Form 1 was found to be fully aligned with the grade 6 mathematics standards and grade 6 Form 2 was found to be acceptably aligned. Two groups independently analyzed grade 6 Form 1 (Tables 7.1 and 7.2). The results were identical for both groups with an acceptable level for all five standards on all four alignment criteria. A few alignment weaknesses were found on grade 6 Form 2. The assessment had an acceptable number of items for each of the five standards, but the DOK levels of the items for Standards I and IV were below the acceptable level. Range was acceptable for all five standards, but balance on Form 2 was weak for Standard IV because of an overemphasis of Objective M.O.6.4.3.

Overall, because grade 6 Form 1 was fully aligned with the mathematics standards no changes are needed. For grade 6 Form 2 to be fully aligned with the mathematics standards, at least three items would need to be replaced. Two items currently targeting content knowledge under Standard I and one item currently targeting content knowledge under Standard IV would need to be replaced with items that have a DOK level that is the same or higher than the assigned objective. If one of the items targeting Objective M.O.6.4.3 is replaced by an item that targets another objective under Standard IV, then this would improve the balance weakness. Reviewers did identify some topics that were not assessed on Form 1. A reviewer from the Grade 6-11 group commented:

There were a number of topics that I expected to see but were not assessed, including understanding the effects of operations with different kinds of numbers, problems using integers, problems using fractions and decimals, problems using rates and proportional reasoning, items where students were to create expressions or equations themselves, problems on coordinate grids, and problems involving measures of central tendency.

Table 7.1
Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 6, Form 1 Standards and Assessments for West Virginia Alignment Analysis 2008 (Grade 6-11 Group)

<i>Grade 6, Form 1 Standards</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>
M.S.6.I - NUMBER AND OPERATIONS	YES	YES	YES	YES
M.S.6.II - ALGEBRA	YES	YES	YES	YES
M.S.6.III - GEOMETRY	YES	YES	YES	YES
M.S.6.IV - MEASUREMENT	YES	YES	YES	YES
M.S.6.V - DATA ANALYSIS AND PROBABILITY	YES	YES	YES	YES

Many of the topics not assessed on grade 6 form 1 also were noted by a reviewer from the Grade 3-6 group:

Under Number and Operations the following were not addressed: Applying the distributive, commutative, associative, and identity properties to numeric expressions; demonstrate an understanding of the effect of multiplying and dividing, whole numbers, fractions, and decimals. Under Geometry the following were not addressed: Apply the concepts of parallel, perpendicular, intersecting, and skew lines to real-world situations; plot polygons on coordinate grids. Under Data Analysis and Probability the following standards were not addressed: identify a real-world situation using statistical measures over time.

Two or three reviewers from the grade 3-6 group did note that some items on grade 6 Form 1 had a source-of-challenge issue (Items 4, 9, 16 and 39). Two reviewers identified Item 38 on Form 2 as having a source of challenge issue.

Table 7.2

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 6, Form 1 Standards and Assessments for West Virginia Alignment Analysis 2008 (Grades 3-6 Group)

Grades 3- 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>
M.S.6.I - NUMBER AND OPERATIONS	YES	YES	YES	YES
M.S.6.II - ALGEBRA	YES	YES	YES	YES
M.S.6.III - GEOMETRY	YES	YES	YES	YES
M.S.6.IV - MEASUREMENT	YES	YES	YES	YES
M.S.6.V - DATA ANALYSIS AND PROBABILITY	YES	YES	YES	YES

Table 7.3

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 6, Form 2 Standards and Assessments for West Virginia Alignment Analysis 2008 (Grade 6-11 Group)

Grade 6, 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>
M.S.6.I - NUMBER AND OPERATIONS	YES	WEAK	YES	YES
M.S.6.II - ALGEBRA	YES	YES	YES	YES
M.S.6.III - GEOMETRY	YES	YES	YES	YES
M.S.6.IV - MEASUREMENT	YES	WEAK	YES	WEAK
M.S.6.V - DATA ANALYSIS AND PROBABILITY	YES	YES	YES	YES

Grade 7

The alignment between the grade 7 mathematics standards and two forms of the WESTEST2 was found to be nearly fully aligned. Only one alignment weakness was found with the DOK levels of items assigned to Standard II on Form 1. The alignment criteria for all other standards on each of the forms were acceptable. Only one item would need to be replaced on Form 1, one item under Standard II, for Form 1 to be fully aligned with the grade 7 standards. Even though the grade 7 test forms and standards were found to be aligned, three reviewers found some grade 7 items that had source-of-challenge

issues (Form 1 Item 39, Form 2 Items 21, 22, 25, and 39). Two reviewers noted for Form 1 topics that were not addressed:

-I expected to see more attention to work with integers, order of operations, work with linear change and proportionality, work with angles, ratio/proportion problems that were not about scale drawings, and measures of central tendency.

-Very weak in Algebra strand for 7th grade. Two problems on translations places too much emphasis on this without any questions on compound figures. More different geometric figures in problems so that the students could use properties of figures to solve. No measures of central tendency problems. More emphasis on proportional reasoning and less on Pythagorean Theorem.

Reviewers noted that Form 2 had too many items on using a scale factor and not enough items on integer computation, the coordinate system, symmetry, and angles.

Table 8.1

Summary of Acceptable Levels on Alignment Criteria for Mathematics 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>
M.S.7.I - NUMBER AND OPERATIONS	YES	YES	YES	YES
M.S.7.II - ALGEBRA	YES	WEAK	YES	YES
M.S.7.III - GEOMETRY	YES	YES	YES	YES
M.S.7.IV - MEASUREMENT	YES	YES	YES	YES
M.S.7.V - DATA ANALYSIS AND PROBABILITY	YES	YES	YES	YES

Table 8.2

Summary of Acceptable Levels on Alignment Criteria for Mathematics 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>
M.S.7.I - NUMBER AND OPERATIONS	YES	YES	YES	YES
M.S.7.II - ALGEBRA	YES	YES	YES	YES
M.S.7.III - GEOMETRY	YES	YES	YES	YES
M.S.7.IV - MEASUREMENT	YES	YES	YES	YES
M.S.7.V - DATA ANALYSIS AND PROBABILITY	YES	YES	YES	YES

Grade 8

For grade 8 the alignment between each of the WESTEST2 assessment forms and the mathematics standards was acceptable. The Categorical Concurrence criterion was acceptable for all five standards with Form 2 along with range and balance. The only alignment issue for Form 1 was a DOK weakness for Standard I. The DOK levels of the Form 1 items assigned to the other four standards were all acceptable. Form 2 had only two minor alignment concerns. Reviewers only found five items that mapped to objectives under Standard IV (Measurement) and only four or five of the 12 items that mapped to objectives under Standard I had a DOK level that was the same or higher than the DOK level of the assigned objective. All of the other alignment criteria and standards were acceptably met.

Only one item on Form 1 and two to three items on Form 2 would need to be replaced to attain full alignment between the assessment forms and the grade 8 standards. At least one Form 1 item and two Form 2 items that mapped to Standard I would need to be replaced by items with DOK levels that are the same or higher than the DOK level of the assigned objective. Also, one item that targeted content under Standard IV would need to be added to Form 2 to attain full alignment.

Table 9.1

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

<i>Grade 8 Form 1 Standards</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>
M.S.8.I - NUMBER AND OPERATIONS	YES	WEAK	YES	YES
M.S.8.II - ALGEBRA	YES	YES	YES	YES
M.S.8.III - GEOMETRY	YES	YES	YES	YES
M.S.8.IV - MEASUREMENT	YES	YES	YES	YES
M.S.8.V - DATA ANALYSIS AND PROBABILITY	YES	YES	YES	YES

The majority of reviewers identified one item on each of the two grade 8 forms as having a source-of-challenge issue—Form 1 Item 23 and Form 2 Item 20. Reviewers did comment that too high of a proportion of items was targeted toward the lower end of complexity of the objectives with only a few items assigned a DOK level 3. Reviewers also noted some topics that were not targeted on the assessment that should be. In particular, the tests had too great of an emphasis on scale factors and not enough on the algebra standard including adding and subtracting polynomials, graphing linear equations with the Cartesian coordinate system by generating a table of values, and real life problems with change over time. Additional suggestions are given in reviewers' notes (Appendix C) and responses to the debriefing questions (Appendix D). One reviewer noted that the grade 8 Form 2 had some seriously flawed items that needed attention

before administering them to students. A reviewer suggested that the items needed to be reviewed by a mathematician.

Table 9.2

Summary of Acceptable Levels on Alignment Criteria for Mathematics 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>
M.S.8.I - NUMBER AND OPERATIONS	YES	NO	YES	YES
M.S.8.II - ALGEBRA	YES	YES	YES	YES
M.S.8.III - GEOMETRY	YES	YES	YES	YES
M.S.8.IV - MEASUREMENT	NO	YES	YES	YES
M.S.8.V - DATA ANALYSIS AND PROBABILITY	YES	YES	YES	YES

Grade 11

The results from the alignment analysis between the two grade 11 assessment forms and the grade 11 mathematics standards indicate that the alignment was acceptable. However, grade 11 Form 1 was the only form that half of the reviewers indicated that the alignment needed major improvement mainly because of poor items. Reviewers noted some items that were flawed such as Form 1 Items 25 and 27. Reviewers did not find the same issues with grade 11 Form 2.

Considering only the minimum acceptable levels for the alignment criteria that have been used in this analysis, the alignment between each of the grade 11 forms and the grade 11 mathematics standards was acceptable. The assessment forms and Standards 1, 2, and 3 were found to be fully aligned. Each form had from eight to 18 items for each of these three standards with an acceptable DOK level, range and balance. The main alignment issue for both forms was an insufficient number of items and items with too low of a DOK level for Standard 4 (Trigonometry). Range and balance for Standard 4 was not an issue because the standard only had two objectives.

Overall, four items would need to be added to grade 11 Form 1 and three items would need to be added to grade 11 Form 2 to attain full alignment. These items would need to target the two objectives under Standard 4 and would need to have an appropriate DOK level. Three reviewers wrote lengthy comments about their concerns with Form 1 indicating that items were too complex:

- When looking at the test as a whole, I am concerned about the accessibility to students and the validity of the results. Because there are so many very dense and complex items, I am concerned that students won't maintain the level of motivation necessary to attempt many of the items or even complete the test.

- Holding students and teachers accountable for their performance on this test will result in information that will not be useful or helpful in informing what students know and can do, the students opportunity to learn, or the instruction that they receive.
- I am very concerned that students at this grade level are not going to be motivated to work these problems. Many of the items are far too complex in context or in the amount of reading or in the mathematics involved in sorting through choices. These items are, in many cases, flawed and need serious attention. Their difficulty stems not from the mathematics but from other item-writing flaws. Many of the other items do not measure what they are really intended to measure. Students are not going to be able to show what they know, so the results are not likely to be valid, and will not be useful to the public.

Reviewers found grade 11 Form 2 to be much better than Form 1. One reviewer did note that more items could be included on Form 2 that related to the big ideas of high school mathematics—functions, multiple representations, and problem solving.

Table 10.1

Summary of Acceptable Levels on Alignment Criteria for Mathematics 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>
M.O.A1 - ALGEBRA I	YES	YES	YES	YES
M.O.A2 - ALGEBRA 2	YES	YES	YES	YES
M.O.G - GEOMETRY	YES	YES	YES	YES
M.O.T - TRIGONOMETRY	NO (2.0)	NO	YES	YES

Table 10.2

Summary of Acceptable Levels on Alignment Criteria for Mathematics 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>
M.O.A1 - ALGEBRA I	YES	YES	YES	YES
M.O.A2 - ALGEBRA 2	YES	YES	YES	YES
M.O.G - GEOMETRY	YES	YES	YES	YES
M.O.T - TRIGONOMETRY	NO (3.0)	NO	YES	YES

TerraNova Grades 3-8 and 11

The TerraNova forms for grades 3-8 and 11 had from 30 to 32 items. The summaries for the alignment criteria for each grade are given in Tables 11.1 through

11.7. The results of the alignment between the TerraNova forms and the West Virginia mathematics standards are very similar across the grades. For each grade 3-8 the TerraNova form had the most items for Standard I (Number and Operations), from 10 to 19 items. Reviewers found at least one item that targeted content under each of the four other standards for grades 3-8. However, the TerraNova forms for grades 4 and 8 were the only forms that had another standard with six items, Standard IV.

The DOK levels of the items on the TerraNova were generally lower than the DOK of the corresponding objective. The items that corresponded to Standard I for grades 4-8, the standard with the greatest number of items, had only 12% to 44% with a DOK level that was the same or above the DOK level of the assigned objective. One to three of the other four standards for each of these grades also had low DOK levels in relationship to the DOK level of the assigned objective. Reviewers frequently noted that the TerraNova items mainly had a DOK level 1 and targeted the lower end of the complexity of the assigned objective.

The range of the items on the TerraNova forms was weak or not acceptable for two or three of the standards for each of grades 3-8. The items were distributed adequately among the objectives that were targeted to have good balance for each of these grades.

Reviewers only found items that mapped to two of the four grade 11 standards—Standard A1 (Algebra) and Standard G (Geometry). Less than 40% of the items for either of these standards had a DOK level that was the same or higher than the DOK level of the assigned objective. Thus, the DOK criterion was not met for any of the four grade 11 standards. Range was acceptable for Standard A1, but not for Standard G. However, nearly half of the items that corresponded to Standard A1 mapped to one objective, M.O.A1.2.2. This resulted in a balance weakness for Standard A1.

Two or more reviewers coded more items on the TerraNova forms to generic objectives than they did on the WESTEST2 forms. Reviewers noted that these items did not fit precisely with the expectations for the grade level and in some cases corresponded to objectives in lower grades. Some of the reviewers' explanations for the grade 6 TerraNova form illustrate their rationale:

- I feel that the breadth (range) of knowledge and the depth of knowledge consistency was extremely low on this assessment. The specific items on the assessment, while they may have fallen under a specific standard, did not match the DOK of the standard. The standards went into more detail than was evident on the test items. This is why on some of the test items I had no other choice but to rank under a generic standard.
- There was not good coverage of the specific three digit standards. There were many items that could only be coded to the general standard. The coverage overall did not go beyond the very general knowledge. This led to many of the specific standards never being addressed by the assessment.
- I found that the majority of these questions were geared to lower grade level objectives, especially in numbers and operations and algebra.

Overall, the alignment between the TerraNova test forms and the West Virginia standards was not very good. The TerraNova test forms targeted mainly Standard I objectives with five or fewer items corresponding to the other four objectives. In general, the TerraNova items had lower DOK levels than the DOK level of the assigned objectives with a number of the items relating to objectives had lower grade levels. Range was also lacking for two or more of the standards for grades 3-8.

Table 11.1

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 3, Form TN, Standards and Assessments for West Virginia Alignment Analysis 2008

<i>Grade 3, Form 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>
M.S.3.I - NUMBER AND OPERATIONS	YES (15.0)	YES	WEAK	YES
M.S.3.II - ALGEBRA	NO (2.0)	YES	NO	YES
M.S.3.III - GEOMETRY	NO (4.0)	YES	YES	YES
M.S.3.IV - MEASUREMENT	NO (5.0)	NO	YES	YES
M.S.3.V - DATA ANALYSIS AND PROBABILITY	NO (4.0)	NO	YES	YES

Table 11.2

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 4, Form TN, Standards and Assessments for West Virginia Alignment Analysis 2008

<i>Grade 4, Form 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>
M.S.4.I - NUMBER AND OPERATIONS	YES (16.0)	WEAK	YES	YES
M.S.4.II - ALGEBRA	NO (1.0)	NT	NT	NT
M.S.4.III - GEOMETRY	NO (4.0)	YES	WEAK	YES
M.S.4.IV - MEASUREMENT	YES (6.0)	NO	YES	YES
M.S.4.V - DATA ANALYSIS AND PROBABILITY	NO (3.0)	NO	YES	YES

Table 11.3

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 5, Form TN, Standards and Assessments for West Virginia Alignment Analysis 2008

Grade 5, Form TN	Alignment Criteria			
Standards	<i>Categorical Concurrence</i>	<i>Depth-of-Knowledge Consistency</i>	<i>Range of Knowledge</i>	<i>Balance of Representation</i>
M.S.5.I - NUMBER AND OPERATIONS	YES (19.0)	NO	YES	WEAK
M.S.5.II - ALGEBRA	NO (2.0)	YES	WEAK	YES
M.S.5.III - GEOMETRY	NO (3.5)	NO	YES	YES
M.S.5.IV - MEASUREMENT	NO (3.0)	YES	NO	YES
M.S.5.V - DATA ANALYSIS AND PROBABILITY	NO (2.5)	YES	NO	YES

Table 11.4

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 6, Form TN, Standards and Assessments for West Virginia Alignment Analysis 2008

Grade 6, Form TN	Alignment Criteria			
Standards	<i>Categorical Concurrence</i>	<i>Depth-of-Knowledge Consistency</i>	<i>Range of Knowledge</i>	<i>Balance of Representation</i>
M.S.6.I - NUMBER AND OPERATIONS	YES (14.0)	NO	NO	WEAK
M.S.6.II - ALGEBRA	NO (5.83)	YES	YES	YES
M.S.6.III - GEOMETRY	NO (3.83)	NO	NO	YES
M.S.6.IV - MEASUREMENT	NO (2.0)	NO	NO	YES
M.S.6.V - DATA ANALYSIS AND PROBABILITY	NO (5.0)	NO	NO	YES

Table 11.5

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

Grade 7, Form TN	Alignment Criteria			
Standards	<i>Categorical Concurrence</i>	<i>Depth-of-Knowledge Consistency</i>	<i>Range of Knowledge</i>	<i>Balance of Representation</i>
M.S.7.I - NUMBER AND OPERATIONS	YES (10.0)	NO	NO	YES
M.S.7.II - ALGEBRA	NO (5.0)	NO	WEAK	YES
M.S.7.III - GEOMETRY	NO (3.0)	YES	WEAK	YES
M.S.7.IV - MEASUREMENT	NO (4.0)	NO	YES	YES
M.S.7.V - DATA ANALYSIS AND PROBABILITY	NO (5.0)	YES	YES	YES

Table 11.6

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 8, Form TN, Standards and Assessments for West Virginia Alignment Analysis 2008

<i>Grade 8, Form 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>
M.S.8.I - NUMBER AND OPERATIONS	YES (12.0)	NO	YES	WEAK
M.S.8.II - ALGEBRA	NO (5.5)	YES	NO	YES
M.S.8.III - GEOMETRY	NO (1.0)	NT	NT	NT
M.S.8.IV - MEASUREMENT	YES (6.5)	NO	YES	WEAK
M.S.8.V - DATA ANALYSIS AND PROBABILITY	NO (4.0)	YES	WEAK	YES

Table 11.7

Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 11, Form TN, Standards and Assessments for West Virginia Alignment Analysis 2008

<i>Grade 11, Form 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>
M.O.A1 - ALGEBRA I	YES (11.5)	NO	YES	WEAK
M.O.A2 - ALGEBRA 2	NO (0.0)	NT	NT	NT
M.O.G - GEOMETRY	YES (6.5)	NO	NO	YES
M.O.T - TRIGONOMETRY	NO (0.0)	NT	NT	NT

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. These comments can be found in Tables (grade).5 and (grade).7 in Appendix C for the WESTEST2 and Appendix F for the TerraNova assessments. Reviewers' comments on the TerraNova items most commonly indicated an item that mapped to an objective in a lower grade. 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 mathematics reviewers' assignment of DOK levels to items was very high for six reviewers for Grades 3-8 and 11 (Table 12). An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the reviewers. The intraclass correlation dropped some when only two

reviewers rated the TerraNova forms. The two reviewers differed greatly in assigning DOK levels to the grade 11 TerraNova items, one coded most of the items as a DOK level 2 while the other coded most of the items as a DOK level 1. A pairwise comparison is used to determine the degree of reliability of reviewers' coding at the objective level and at the learning goal level. The standard pairwise comparison values are high and along with those for the objective pairwise comparison. 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 reviewer agreement after some discussion.

Table 12
Intraclass and Pairwise Comparisons, West Virginia Alignment Analysis for Mathematics Grades 3-8 and 11 Assessments

Grade	Intraclass Correlation	Pairwise Comparison:	Pairwise: Objective	Pairwise: Standard
3 Form 1	.92	.80	.88	.96
3 Form 2	.92	.83	.88	.96
4 Form 1	.94	.82	.91	.96
4 Form 2	.96	.90	.88	.91
5 Form 1	.94	.84	.91	.94
5 Form 2	.92	.84	.81	.90
6 Form 1	.98	.97	.96	.96
6 Form 2	.99	.98	.98	.99
6 Form 1 by Grade 3-6 Group	.96	.91	.82	.92
7 Form 1	.98	.98	.92	.94
7 Form 2	.95	.86	.92	.97
8 Form 1	.91	.90	.94	.96
8 Form 2	.92	.90	.90	.92
11 Form 1	.88	.79	.92	.94
11 Form 2	.90	.84	.98	1.00
3 Form TN (N= 2)	1.00	1.00	1.00	1.00
4 Form TN (N= 2)	.78	.90	.90	1.00
5 Form TN (N= 2)	.90	.93	.86	.96
6 Form TN (N= 6) by Grade 3-6 Group	.66	.64	.62	.82
7 Form TN (N= 2)	.72	.75	1.00	1.00
8 Form TN (N= 2)	.79	.84	.93	.93
11 Form TN (N= 2)	.28	.52	.94	.94

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 mathematics 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 these documents for grades 6-8 and 11. Both groups independently analyzed the alignment with grade 6 Form 1 and attained the same results. Half of the reviewers were from West Virginia and half were from other states. The reviewers included mathematics education content experts, district mathematics supervisors, mathematics teachers, and mathematics assessment experts. Two forms of the WESTEST2 assessment and one TerraNova form were analyzed for each grade.

The alignment for each of the WESTEST2 forms with the West Virginia 21st century mathematics content standards for grades 3-8 and 11 was at least acceptable with only minor changes needed to attain full alignment (see summary table below). Three WESTEST2 forms were fully aligned with the 21st century standards—grade 3 Form 1, grade 6 Form 1, and grade 7 Form 2. The main alignment deficiency for grades 3-5 was fewer than six items, the acceptable level for the Categorical Concurrence criterion, for either or both Standards II (Algebra) and V (Data Analysis and Probability). The depth-of-knowledge (DOK) levels of the items, except for grade 5 Form 2, were suitable along with the range. One or two objectives had a greater proportion of items than other objectives for one standard for grade 3 Forms 1 and 2 and grade 4 Form 1. These balance weaknesses were considered more a matter of preference since the other alignment criteria for the standard were acceptable.

For grades 6-8, the main alignment concern was with the DOK levels of items for one or two standards—Standard I (Number and Operations) for grade 6 Form 2 and two grade 8 forms, Standard II (Algebra) grade 7 Form 1, and Standard IV (Measurement) grade 6 Form 2. Otherwise the assessment forms had a sufficient number of items for each standard (except Standard IV grade 8 Form 2), range, and balance.

Reviewers found an insufficient number of items on both of the grade 11 forms for the trigonometry standard, only two or three items. These items also were lower in complexity than the assigned objectives. Range and balance both were suitable.

Overall, up to four 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, particularly grade 11 Form 1.

The TerraNova forms for all seven grades analyzed were not aligned with the West Virginia 21st century mathematics standards. The TerraNova forms for grades 3-8 had at least some items for each of the five mathematics standards. From one-third (grade 7) to two-thirds (grade 5) of the items on each form mapped to Standard I (Number and Operations). The other four standards generally had less than six items for each of the

grades. The DOK levels of the items were low for two to four standards for each grade. Range was poor for two to four of the standards for each grade. Balance for Standard I for grades 5, 6, and 8 was weak because one objective was over emphasized.

Reviewers did not find any items on the grade 11 TerraNova form that targeted the Algebra II or trigonometry standards. The DOK levels of the items that targeted the other two standards were low. Range for the Algebra I standard was barely acceptable and not acceptable for the geometry standard.

Summary Table

Percent of West Virginia Mathematics 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	100	80	0
3 Form 2	60	100	100	80	2
4 Form 1	60	100	100	80	2
4 Form 2	80	100	100	100	1
5 Form 1	60	100	100	100	3
5 Form 2	80	80	100	100	3
6 Form 1	100	100	100	100	0
6 Form 2	100	60	100	80	3
7 Form 1	100	80	100	100	1
7 Form 2	100	100	100	100	0
8 Form 1	100	80	100	100	1
8 Form 2	80	80	100	100	2-3
11 Form 1	75	75	100	100	4
11 Form 2	75	75	100	100	3

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 science education*. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison: University of Wisconsin, Wisconsin Center for Education Research.