



**An Alignment Analysis of
Washington State's College Readiness
Mathematics Standards
With Various Local Placement Tests**

Big Bend Community College
Washington Public Universities
Whatcom Community College

June 5, 2006

TABLE OF CONTENTS

Introduction.....	3
Analysis Protocol.....	4
Major Findings & Recommendations.....	8
Appendix A: Alignment Analysis for Big Bend Community College	10
Appendix B: Alignment Analysis for Washington Public Universities	18
Appendix C: Alignment Analysis for Whatcom Community College	25
Appendix D: Biographies	34

INTRODUCTION

For the past two years, Achieve, Inc. has worked with the Transition Mathematics Project (TMP), a collaborative project of K-12 schools, community and technical colleges, and baccalaureate institutions within the state of Washington, as it developed mathematics standards that “define the core knowledge and skills expected of students entering college-level mathematics courses and courses with quantitative components.” As part of its work with the TMP, Achieve was asked to assess the alignment between a sample of college placement tests used by Washington’s state postsecondary institutions and the College Readiness Mathematics Standards (CRMS) recently completed by the TMP.

Three sets of placement tests were submitted for alignment analysis. The four-year institutions—Central Washington University, Eastern Washington University, Western Washington University, the Washington State University System, and the University of Washington—submitted two placement tests, the Intermediate Mathematics Placement Test (IMPT) and the Advanced Mathematics Placement Test (AMPT). Big Bend Community College submitted two levels of their mathematics placement test, Form 1A and Form 2A. Whatcom Community College submitted four mathematics placement tests: Basic Mathematics, Beginning Algebra, Intermediate Algebra, and Pre-Calculus. It is important to note that these placement tests were all developed independently of the TMP and have been used by their various institutions for several years prior to the development of the CRMS.

The CRMS describe what students should know and be able to do in three process standards—*Problem-Solving and Reasoning*, *Communication* and *Connections*—and five content standards—*Number Sense*, *Geometry*, *Statistics and Probability*, *Algebra*, and *Functions*. Each standard includes “a set of *components* which represent specific elements within the broader standard. For each component there is *evidence of learning indicators* which clarify specific behaviors or performances that would demonstrate the particular component involved.” Extra expectations, which are noted in italics, are embedded throughout the content standards, identifying knowledge and skills needed by students intending to take higher-level mathematics courses (e.g., pre-calculus and/or calculus) when they enter college. Test items were mapped by Achieve staff to the CRMS using the greatest degree of specificity possible—that is, to an appropriate *learning indicator* wherever possible. In some cases, an item appeared to clearly fall within the intent of a particular component but could not be aligned with one of the learning indicators; in such a case, the item was mapped to the component itself.

ANALYSIS PROTOCOL

A review team comprised of two Achieve consultants confirmed the mapping of the items for each test, took each test, and then rated each item against the component or learning indicator to which it was mapped using four criteria—*Content Centrality*, *Performance Centrality*, *Source of Challenge*, and *Level of Demand*.

Content Centrality analyzes the degree to which the main content knowledge measured by the item is consistent with the most important content that students are expected to know as expressed in the targeted objective. *Performance Centrality* analyzes the consistency of the type of performance demanded by the item to the performance described in the targeted objective. For both content and performance, a rating of 2 indicates that the content or performance is clearly consistent. A rating of *1a* indicates that the objective is too general in its statement of core content or performance to be assured of an item's strong alignment. While the CRMS may contain other general objectives, there were only three to which items were mapped that were considered too general in terms of content to be confident of alignment. All items mapped to these objectives received a Content Centrality rating of *1a*.

- Component 1.3: *Use logical reasoning and mathematical knowledge to obtain and justify correct solutions*
- Learning Indicator 3.2.a: *Connect mathematical definitions and procedures with underlying math concepts.*
- Learning Indicator 7.1.f: *Use appropriate properties to simplify an expression, resulting in an equivalent expression.*

Learning Indicator 7.1.f was considered too general in performance as well as in content to be confident of alignment; therefore, items mapped to 7.1.f received a *1a* for both Content Centrality and Performance Centrality. Items mapped to two additional objectives were rated *1a* for Performance Centrality only.

- Component 7.5: *(extra expectations) Demonstrate an understanding of sequences and series.*
- Learning Indicator 8.4.c: *Abstract mathematical models from word problems and interpret solutions in the context of these source problems.*

A rating of *1b* indicates that an item assesses part, and the less central part, of a compound objective. Ratings of *1b* for Content Centrality were common across the tests. This was often a reflection of the compound nature of many of the objectives. For example, learning objective 6.3.a could align with items that address a variety of statistical measures.

- Learning Indicator 6.3.a: *Use measures of central tendency (mean, median, mode) and spread (range, quartiles) to summarize data, draw inferences, make predictions, and justify conclusions.*

A second reason why an item might receive a Content Centrality rating of *Ib* is that the item involves only the simplest aspect of content (e.g., type of number, function, or expression) included in the objective. For example, items mapped to Learning Indicator 4.2.a received a rating of *Ib* for Content Centrality if the computations they called for involved only whole number arithmetic when the objective is written about *real* numbers.

- Learning Indicator 4.2.a: *Complete multi-step computations using the order of operations and the properties of operations (associative, commutative, distributive, etc.) in situations involving combinations of real numbers.*

Similarly, items mapped to objectives that call for students to carry out multiple actions often receive a rating of *Ib* for Performance Centrality. This occurs, for example, when items align with the least cognitively demanding of the performances called for in an objective. Learning Indicator 6.3.a above is an example of an objective where multiple actions are called for. Other CRMS objectives ask students to “find and verify,” “apply and justify,” or “draw and justify [conclusions]” whereas the items examined rarely required the application of the more difficult of these performances.

A rating of *Ic* for Content Centrality or for Performance Centrality indicates that the objective is too specific to be confident of strong alignment. For example, two items (one on each of two different tests) were mapped to Learning Indicator 7.2.d.

- Learning Indicator 7.2.d: *Simplify quotients of polynomials given in factored form, or in a form which can be factored (extra expectations) or determine if irreducible over the real numbers.*

Both of the items that mapped to this Learning Indicator call for students to determine the remainder when a polynomial is divided by a linear binomial. While this Learning Indicator is the closest match to the content—polynomial quotients—the rest of the descriptor in the objective does not clearly match the item. These two items received a *Ic* for Content Centrality. For Performance Centrality, only three items across all tests received a *Ic* rating. These items were mapped to Learning Indicator 8.1.c or Learning Indicator 8.2.a, both of which indicate that the purpose of the performance is to apply it to graphing. Since none of the three identified items involved a graph, it was difficult to confirm a strong alignment.

- Learning Indicator 8.1.c: *Understand and interpret function notation, particularly as it relates to graphic displays of data.*
- Learning Indicator 8.2.a: *Evaluate functions to generate a graph.*

A *Source of Challenge* issue is raised when the most demanding thing about an item is *not* the mathematics that is targeted. Three items across the eight tests submitted for review were identified as having an inappropriate Source of Challenge and were given a rating of *0* in this category. In one case, two of the five distracters were algebraically equivalent, and either could have been selected as a correct answer. A second item involved a loan for which no interest was added. The third asked students to find the

number of items having a given characteristic when the answers offered were the percentages of items. Four additional items were noted by reviewers as having minor problems, but the level of concern was not such that these items were identified as having inappropriate sources of challenge.

Finally, each item was given a rating for *Level of Demand*. There are four levels of cognitive demand described in the Achieve Assessment-to-Standards Protocol; however, it would be difficult, if not impossible, for a multiple-choice item to require reasoning that would be considered Level 4. While the four levels are defined below, for purposes of this analysis only three levels were used.

- Level 1: Recall of a fact, information or procedure;
- Level 2: Application of a skill or concept/use of information or conceptual knowledge;
- Level 3: Strategic thinking that requires reasoning, developing a plan, working with some complexity and considering more than one possible approach or answer;
- Level 4: Extended thinking that requires an investigation and time to think and process multiple conditions of the problem.

Subsequent to this item-level analysis, the lead reviewer made a more holistic evaluation of the collection of items on each test with respect to three characteristics: *Balance, Range, and Level of Challenge*. *Balance* is a measure of how well a set of test items mapped to a standard reflects the emphases that a standard and its related objectives give to particular content. If a test is designed to assess a set of standards, the balance of the items should reflect the balance of content knowledge and skills expressed in the standards. As is the case with the majority of college and work readiness standards Achieve has had an opportunity to analyze, the CRMS are strongly weighted toward algebraic concepts, which are valued as predictors of college-level success in future mathematics courses. Approximately 48 % of the components and learning indicators in the CRMS are in the Algebra and Functions standards. Despite this emphasis, the TMP’s standards reflect an excellent breadth of mathematical content and process knowledge, better breadth and balance than any of the eight tests submitted for review.

During the analysis of each of the eight tests, the balance of the test was determined based on how well the set of items reflected the emphases defined in the CRMS. The table below shows the distribution of the 171 components and learning indicators contained in the CRMS across the eight standards. The number of objectives in each CRMS standard that are partly or wholly *extra expectations* is given in parentheses and italics.

CRMS Standard	Number (%) of Objectives
Reasoning/Problem-Solving	13/171 (<i>0</i>) (~8%)
Communication	11/171 (<i>0</i>) (~6%)
Connections	14/171 (<i>0</i>) (~8%)
Number Sense	16/171 (<i>3</i>) (~9%)

Geometry	19/171 (4)	(~11%)
Statistics and Probability	16/171 (0)	(~9%)
Algebra	42/171 (24)	(~25%)
Functions	40/171 (26)	(~23%)

Range, in contrast to balance, is a measure of coverage. For this criterion, all items that map to each of the standards are considered as a set. Range is the fraction of total objectives in a standard that are assessed by at least one item. An analysis of a test's *Level of Challenge* is similarly based on all of the items mapped to a given standard and assesses the range of levels of cognitive demand they represent. A good test includes both fairly simple and more complex tasks within each standard.

MAJOR FINDINGS & RECOMMENDATIONS

Washington has done a substantive amount of work in its efforts to define the mathematical knowledge and skills students need to be prepared upon high school graduation for entry into credit-bearing mathematics courses in college. The CRMS developed by the TMP, coupled with the state’s Grade Level Expectations—which are expected to be extended to encompass all four years of high school—set a strong foundation for students. Standards alone are not sufficient to ensure student preparedness for postsecondary education, and it is important that key aspects of the K-12 and postsecondary communities are aligned if students are to get consistent messages about what is important in mathematics. With this in mind, the TMP commissioned this analysis of three sets of postsecondary course placement tests to determine the degree of alignment between its articulated expectations and the *de facto* expectations students encounter when they go to enroll in their first college mathematics course. Achieve’s major findings from this analysis are as follows:

- **The CRMS developed by the TMP partnership involving K-12 educators, postsecondary educators, and the business community suggest a broader and richer view of what is needed for future success in postsecondary mathematics than is defined in the placement tests used to place students into entry-level college mathematics courses.**

The CRMS place considerable emphasis not only on the content students need to know to be successful in college-level mathematics classes, but also on the skills and processes they need. The fact that the three mathematical process standards—Reasoning/Problem-Solving, Communication, and Connections—precede the content-specific standards shows the level of importance that the TMP partners attach to such skills. None of the placement tests examined as part of this study include more than a few items that tap into these mathematical processes. Rather, the items, and particularly those on the community college assessments, tend to require computation and symbolic manipulation, which are important foundational skills, but do not exemplify what it means to be able to do meaningful mathematics. In addition, the CRMS define a broad set of content standards that encompass all of the domains of mathematics typically addressed in middle school and high school—Number Sense, Geometry, Probability/Statistics, Algebra, and Functions. While both the CRMS and the placement tests tend to place their greatest emphasis on Algebra and Functions—which are important precursors to college-level mathematics—the treatment of these two domains of mathematics is balanced in the CRMS with a relatively balanced treatment of Number Sense, Geometry, and Probability/Statistics. Comparable balance was not evidenced in the placement tests. In particular, the community college assessments tend to place heavy emphasis on Algebra, Functions, and Number Sense.

- **Overall, the placement tests examined as part of this study do an insufficient job of assessing the CRMS developed by the TMP.**

Coverage of the standards is weak, but this is not totally unexpected given that the tests now in use were developed long before the CRMS came into existence. Refinements and revisions to placement tests now in place can remedy this by taking the CRMS into consideration as assessment items are developed. In particular, the inclusion of items that provide students with opportunities to display their reasoning and problem-solving skills would improve alignment between the tests and the CRMS. Additionally, rather than repeatedly assessing those less challenging aspects that might be easier to assess, items that more fully assess the breadth of the Algebra and Functions expectations and other items that more fully assess the level of expectation defined in the CRMS for Number Sense, Geometry, and Probability/Statistics would improve alignment. Given the number of expectations defined in the CRMS, it would be difficult to develop tests that cover everything, but steps can be taken to do a better job of covering the standards. *(Please see Appendices A, B and C for a detailed description of findings regarding specific tests.)*

- **Some components and learning indicators in the CRMS are written in ways that made it difficult to tell whether test items are entirely consistent with the articulated expectations.**

Use of the CRMS in this alignment analysis uncovered some aspects of the standards that could be refined. In a few instances (detailed in the institution-specific reports, as applicable), components or learning indicators are written so broadly and in such general terms that it is difficult to tell whether a particular item maps well to that objective. This happened most often at the component level and affected items that could only be mapped at that level. Other components or learning indicators were too specific, meaning that items that related closely to the content and performance defined in the expectation did not fit clearly with some part of the standard. Other components and learning objectives contained several content topics or performance verbs joined by the conjunction “and,” which implies that students should be able to do all aspects of the standard. While this is not necessarily a bad thing, it is important when developing assessments to align with the standards that care be taken to assess multiple content and performance aspects of the standard rather than repeatedly assessing one aspect. This packaging of the CRMS will be a challenge to those tasked with developing assessment items to measure them.

APPENDIX A: BIG BEND COMMUNITY COLLEGE

Big Bend Community College submitted two placement tests for this alignment analysis. Test Form 1A is given to incoming students who have taken less than two years of high school algebra. Test Form 2A is used for incoming students who have had two years or more of high school algebra. All test items are presented in a multiple-choice format having five possible answers, the fifth of which is always “none of these.” The two tests have thirty-five common items (items #1 – 35 on Form 2A are identical to items #14 – 17 and #25 – 55 on Form 1A). Some graphs and/or data tables included with this report will show the breakdown for the common items and for each set of items unique to a particular test form. Both unique and common items have been included in the analysis and summary statistics for each of the two test forms.

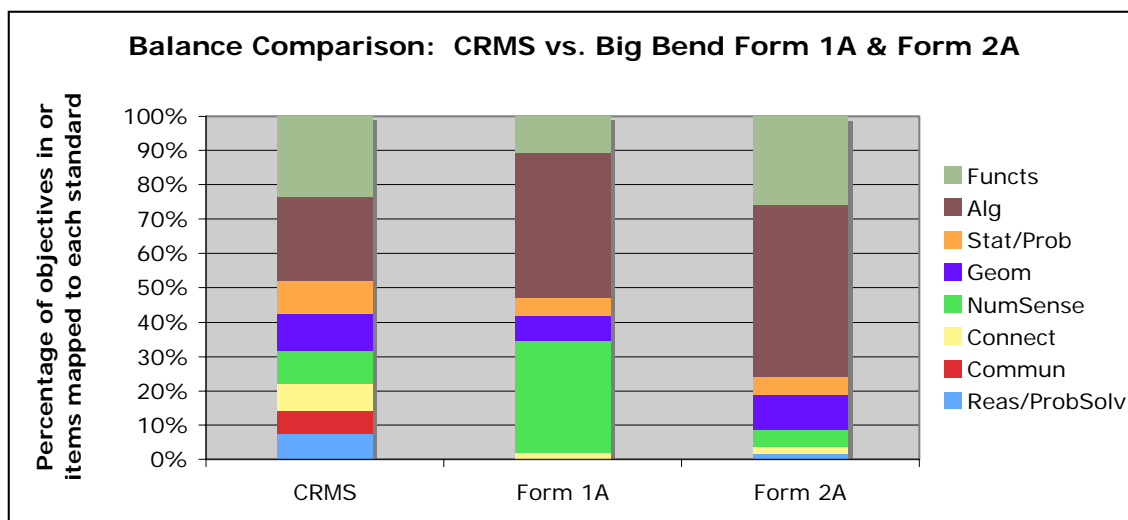
Map of Test Items to the College Readiness Mathematics Standards

Test Form 1A has a total of 55 items while Form 2A has a total of 59 items, 24 of which are unique to the test form. Seventy-eight of the 79 unique items across the two test forms were able to be mapped to expectations of the CRMS. One item, #57 on Form 2A, could not be mapped to any of the CRMS process or content standards since it deals with the linear combination of two vectors, a topic not addressed by the CRMS. Whenever possible, items were mapped to specific learning indicators. Twelve items unique to Form 1A, three unique to Form 2A, and two common items did not align with any specific learning indicator and were, therefore, mapped to the more general component level. Seven (~13%) of the items on Form 1A and 19 (~33%) of the items on Form 2A mapped to *extra expectations* of the CRMS—expectations defined for students intending to enter college taking pre-calculus or calculus. While this seems a bit high, especially on Form 1A which is intended to place students into the lowest level credit-bearing courses, all of the *extra expectations* mappings occurred in Algebra and Functions where the preponderance of the *extra expectations* objectives occur in the CRMS.

Balance

Test Form 1A places a definite emphasis on Number Sense and Algebra. Out of 55 items on the test, 18 (or ~33%) mapped to Number Sense and 23 (or ~42%) mapped to Algebra while an additional six items (or ~11%) mapping to Functions, which is a content area closely related to Algebra. The heavy emphasis on computation found in Form 1A is not reflected in Form 2A where only three items (~5%) common to Form 1A map to Number Sense. In the more advanced test, the emphasis shifts more heavily to Algebra (29 of the 58 mapped items or 50%) and Functions (14 of the 58 mapped items or ~24%).

The graph below illustrates the balance of the 55 items on Form 1A and the 58 items on Form 2A that could be mapped to either learning indicators or components of the CRMS across all process and content standards and compares the balance of these two tests to the balance of the standards themselves.



Heavy emphasis on computation, algebra and function is understandable in light of the content of many entry-level college mathematics courses and in light of the demands of many quantitative courses in other discipline areas. It should be noted, however, that this emphasis is significantly greater in both forms of the Big Bend placement tests than it is for the CRMS—75% of the items on Form 1A assess Number Sense or Algebra and 50% of the items on Form 2A are mapped to Algebra while only 5% map to Number Sense. The CRMS, on the other hand, have slightly less than 35% of the objectives assigned to either the Algebra or Number Sense standards. In addition, these two tests do not reflect the breadth of the standards developed by the TMP that emphasize processes skills—Problem-Solving, Communication, and Connections—as important components in mathematical knowledge and that include a stronger emphasis on Statistics and Probability.

Balance within each standard is also an area of importance. Of the 18 test items on Form 1A that mapped to Number Sense, 11 of them were mapped to Component 4.2:

- *Accurately and efficiently compute with real numbers in all forms, including rational exponents and scientific notation.*

An additional 5 were mapped to Learning Indicator 4.2.a:

- *Complete multi-step computations using the order of operations and the properties of operations (associative, commutative, distributive, etc.) in situations involving combinations of real numbers.*

Items mapped to Component 4.2 are those that involve only one-step computations. This provides poor balance across the Number Sense standard. In addition, ten items mapped to Component 4.2, and one item mapped to Learning Indicator 4.2.a involves computations of whole numbers only. This does not provide good balance across the richness of number types indicated in both the component and learning indicator. Other

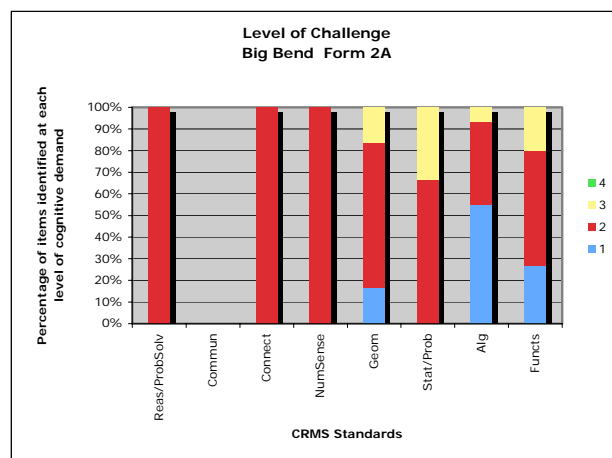
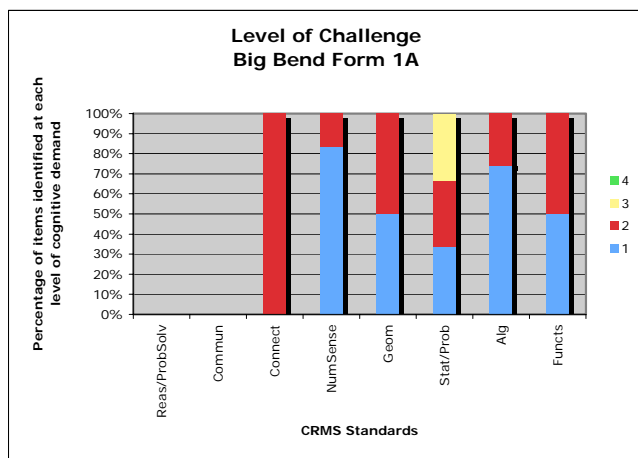
examples where balance within a standard on Form 1A could be improved are indicated below:

- Two of the three items mapped to Geometry (both of the items mapped to Learning Indicator 5.3.b) test proportional reasoning.
- Both items mapped to Learning Indicator 6.3.a deal with mean.

On the other hand, the items on Form 1A that were mapped to Algebra learning indicators exhibit good balance across the variety of types of problems mentioned in the objectives. Form 2A exhibits good balance within each of the standards.

Level of Challenge

Collectively, the items from the two Big Bend tests examined as part of this study exemplify a range of levels of cognitive demand. However, there is a difference between tests, and within each test, some standards tend to be assessed with more cognitively demanding items than other standards. The graphs below show that the Level of Challenge is greater in Form 2A than in Form 1A. This is to be expected since Form 2A is intended to identify students who may be ready for a higher entry-level mathematics course than students taking Form 1A.



There were only two items mapped to the process standards—one common to both tests mapped to Connections and one unique to Form 2A mapped to Reasoning and Problem-Solving. Both of these items received Level 2 rating of cognitive demand. The components and learning indicators found in the process standards describe the type of performance that invites deeper thinking that might result in a Level 3 rating, but with only two items assessing mathematical processes, it is difficult to expect any range of levels.

Except for the three items mapped to the Number Sense standard on Form 2A—all of which were items common to both tests and all of which were rated Level 2—a mix of levels of cognitive demand is exhibited by the item sets mapping to all of the CRMS

content standards. Because they include Level 3 items, the item sets aligning with the Statistics and Probability standard from Form 1A and the Geometry, Algebra, and Functions standards from Form 2A show the most balanced distribution with respect to Level of Challenge.

Range

Range is a measure of coverage and is measured by the fraction of total objectives (components and learning indicators) within a given standard that are assessed by at least one item. The table below illustrates the range of coverage for each CRMS standard on each of the two Big Bend tests. The number of items that mapped to expectations designated as “extra expectations” (those needed by students taking higher level mathematics courses when they enter college) are indicated in italics.

CRMS Standard	Portion of Standards Assessed by Form 1A		Portion of Standards Assessed by Form 2A	
Reasoning/Problem-Solving	0/13	(0%)	1/13	(8%)
Communication	0/11	(0%)	0/11	(0%)
Connections	1/14	(7%)	1/14	(7%)
Number Sense	3/16	(19%)	3/16	(19%)
Geometry	3/19	(16%)	5/19	(26%)
Statistics and Probability	2/16	(13%)	3/16	(19%)
Algebra	15/42	(36%)	21/42	(50%)
Functions	5/40	(13%)	14/40	(35%)

Generally, when tests are constructed to align with standards, range values at or above 67% are considered good, while those from 50% to 66% are considered acceptable. Using this as a standard, Big Bend Form 1A never reaches an acceptable range of coverage for any standard. Form 2A barely meets an acceptable level only for Algebra with 50% coverage. Low levels of coverage demonstrate a fairly significant mismatch between the tests and the TMP’s CRMS standards. In the case of the Big Bend tests, this low level of coverage is often the result of either very few items mapped to a given standard—this is the case for Geometry and Statistics and Probability on both tests—or the result of having multiple items mapped to the same objective—this is definitely the case for Number Sense on Form 1A and, to a lesser extent, is an issue on Form 2A within Algebra. It is important to keep in mind that with only 78 unique items on the Big Bend tests and 171 CRMS objectives, excellent coverage would be difficult to obtain.

Source of Challenge

Occasionally, an item contains some aspect that may provide a challenge to students that comes from something other than the targeted mathematics content. While no items on the Big Bend tests were identified as definitively having inappropriate sources of challenge, two items common to both tests and two additional items unique to Form 2A were identified by reviewers as somewhat problematic.

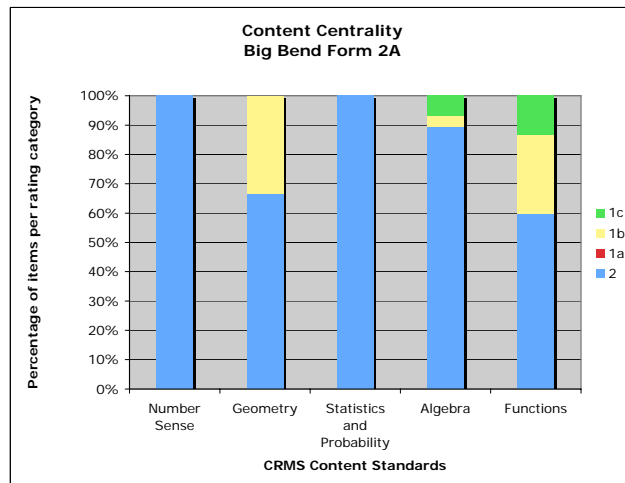
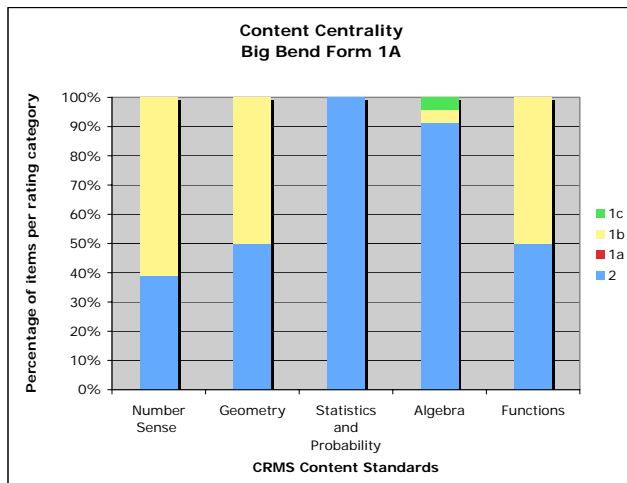
- #49 on Form 1A (also #49 on Form 2A) mapped to Component 5.3 (*Use properties of and relationships between 2-D or 3-D figures to draw and justify conclusions about a situation represented with such figures with or without a coordinate system.*). The item asks for the dimensions of a square but only two of the four answers provided as distracters are dimensions of a square.
- #51 on Form 1A (also # 31 on Form 2A) mapped to Learning Indicator 7.1.g (*Recognize the equivalence between expressions with rational exponents and radicals.*). While one answer choice is presented in a generally preferred “simplest” form with the negative removed from the radical, students could demonstrate significant understanding of simplifying a cube root expression and miss the answer by selecting choice (d) that leaves the negative under the radical.
- #45 on Form 2A mapped to *extra expectations* 8.2.e (*Given an algebraic representation of a rational function, find the intercepts, asymptotes (horizontal, vertical, and slant), and holes (discontinuities), then sketch the graph.*). Students could confuse the definition of horizontal and vertical and miss the problem while understanding much of the core of the learning indicator.
- #56 on Form 2A mapped to *extra expectations* 8.6.h (*Generate algebraic and graphical representations of inverse trig functions (arcsin, arccos, arctan), and determine domain and range*). The answers presented all include $\sqrt{3}$ while the correct answer involves $\sqrt{13}$. This was likely unintended; however, it makes it easy for students to see that “none of these” is the only possible choice.

Items identified as having an inappropriate Source of Challenge are not used in analyses with respect to Content Centrality, Performance Centrality or Level of Demand. Because no items on Form 1A or 2A were so identified, all items from the Big Bend tests are used in the following analyses.

Content Centrality

The review team assigned ratings to the two items (#33 Form 1A/#13 Form 2A mapped to Learning Indicator 3.2.a and #51 Form 1A mapped to Component 1.3) that mapped to process standards even though they are not strictly speaking mathematical content. In both cases, the items received a rating of *1a* indicating that the objective is too general to be assured of an item’s strong alignment.

As can be seen in the graphs that follow, the strength of content alignment varies by test and by content domain. Ratings of *2* indicate consistent alignment, ratings of *1b* indicate that the item assesses only part—and the less central part—of a compound objective, and a rating of *1c* indicates that the objective is so specific that the mapping is only partially consistent. Ratings of *1a* are not reflected in these graphs since they only arose in conjunction with items mapping to the CRMS process standards.



Where ratings for Content Centrality are not consistent, part of the explanation might be that these tests were not written to reflect the standards to which they are being aligned. Having these standards should provide guidance as future test development is undertaken. A second explanation for ratings that are not consistent, in some areas, has to do with the standards themselves. Some components and learning indicators are written to encompass several content topics joined by the conjunctive (“and”). Learning Indicator 5.3.b is a good example of this:

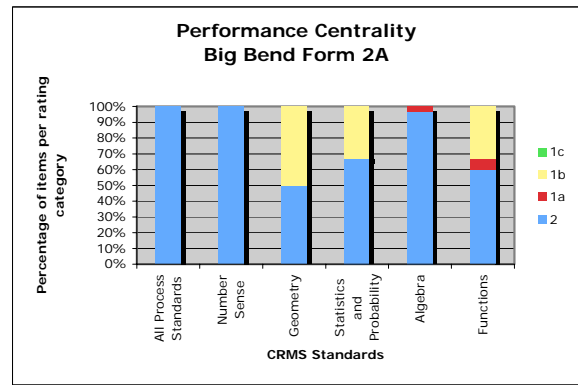
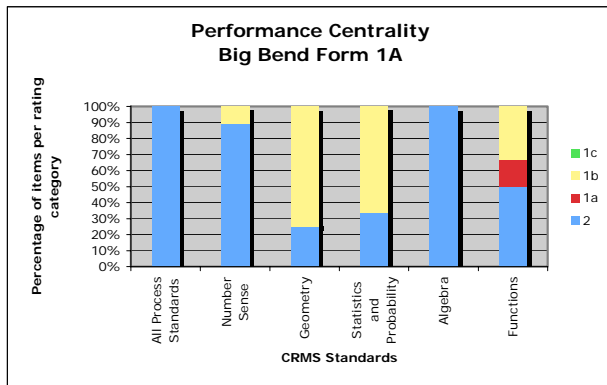
- *Apply and justify the applicability of transformations, congruence, similarity, ratios, and proportions in problem-solving situations.*

It would be difficult for any test item to assess more than one or two of the content topics in a single question. It is even more difficult to say that an item hitting only one area is hitting the “central part” of the objective. One way that a test can address this issue is to have several questions mapping to the same compound objective and thereby assessing a different content topic for each question. While some of the ratings might still be *1b*, they would at least assess the breadth of knowledge indicated by the objective.

Performance Centrality

Reviewers rated both of the Big Bend tests as being strong with respect to Performance Centrality. The ratings for Performance Centrality are analogous to those used for Content Centrality (2, *1b*, *1a*, and *1c*) and have comparable meanings. Forty-six of the 55 items (~85%) on Form 1A received a rating of 2. Forty-eight of the 58 mapped items (~83%) on Form 2A were rated 2 for Performance Centrality. Most of the remaining items received a *1b* rating. Generally, the rating of *1b* was assigned to items that mapped to standards having multiple verbs joined by “and” (e.g., “apply and justify,” “find and simplify”). In these cases, the items generally only addressed the more basic of the two performances. A high percentage of the items mapped to the Geometry and Statistics and Probability standards on Form 1A were rated *1b*. It is important to keep in mind that when working with low numbers of items, percentages can be misleading. Only five items in total mapped to these two standards combined.

The graphs below show the distribution of Performance Centrality ratings for the two Big Bend placement tests. Process standards are included in analysis of performance but are combined here since so few items mapped to them. As was the case with Content Centrality, ratings of 2 indicate consistent alignment, ratings of *1b* indicate that the item assesses only part—and the less cognitively demanding part—of a compound objective, and a rating of *1a* indicates that the objective in the CRMS to which an item maps is too general to be assured of the item’s strong alignment. No items on the Big Bend tests were rated *1c*, which would have indicated that the objective to which that item mapped is so specific that the mapping is only partially consistent.



Level of Demand

Perhaps the area for greatest concern is the level of cognitive demand required of students taking the Big Bend tests. Achieve’s rating scale defines four levels of cognitive demand although a multiple-choice format generally only allows for three to be assigned. Level 3 items are the most cognitively demanding of the three types of items, requiring that students use reasoning and planning skills and employ a higher level of thinking than the other two levels. Level 2 items require students to make some decisions as to how to approach a problem, whereas Level 1 items require students to recall information or perform a well-known algorithm or procedure. The Big Bend tests assess a fairly low level of cognitive demand for instruments that are used to guide the course placement of students entering college.

Ninety-six percent of the items on Form 1A received a Level 1 or 2 rating for Level of Demand. Of the two problems rated Level 3, one (#49) presents two answer choices that do not fit the conditions of the problem (as noted under the Source of Challenge section above), making the answer relatively easy to guess. Only one problem (#26) on the entire test really requires reasoning or planning skills and offers a reasonable choice of answers. The more advanced test, Form 2A, is only slightly more demanding. Seven of the 58 mapped items (~12%) received a Level 3 rating. One of these (#29) was identical to the Form 1A item that had limited correct responses, two of them (#40 and #50) were only marginally assigned a Level 3 rating by reviewers, and two of them addressed

learning indicators that were identified as “extra expectations” (#54 and #58). Just over half of the items on Form 2A received a Level 2 rating, and the rest were deemed to be recall or the application of a rote procedure and were rated Level 1 for Level of Demand.

On both tests, the emphasis is on assessing student skills. This will provide information that is helpful in some ways but gives little insight into the learning capabilities of entering students. The TMP’s CRMS appear to suggest a broader and richer view of the kind of thinking needed for future success in postsecondary mathematics.

APPENDIX B: WASHINGTON PUBLIC UNIVERSITIES

Two mathematics placement tests, The Intermediate Mathematics Placement Test (IMPT) and the Advanced Mathematics Placement Test (AMPT), collectively used by Washington's four-year colleges and universities were submitted for alignment. Each of the five schools that use these tests—Central Washington University, Eastern Washington University, the University of Washington, Washington State University and Western Washington University—offer a choice of the two tests to students and have developed individual cut scores for each test that identify appropriate placement for students that attend that school. Some of the schools require a particular score on the AMPT to be placed into certain advanced mathematics courses including calculus.

Map of Test Items to the College Readiness Mathematics Standards

The IMPT consists of 35 items, 33 of which could be mapped to specific learning indicators of the CRMS. The remaining two items—#21 and #22—were both mapped to Component 8.2:

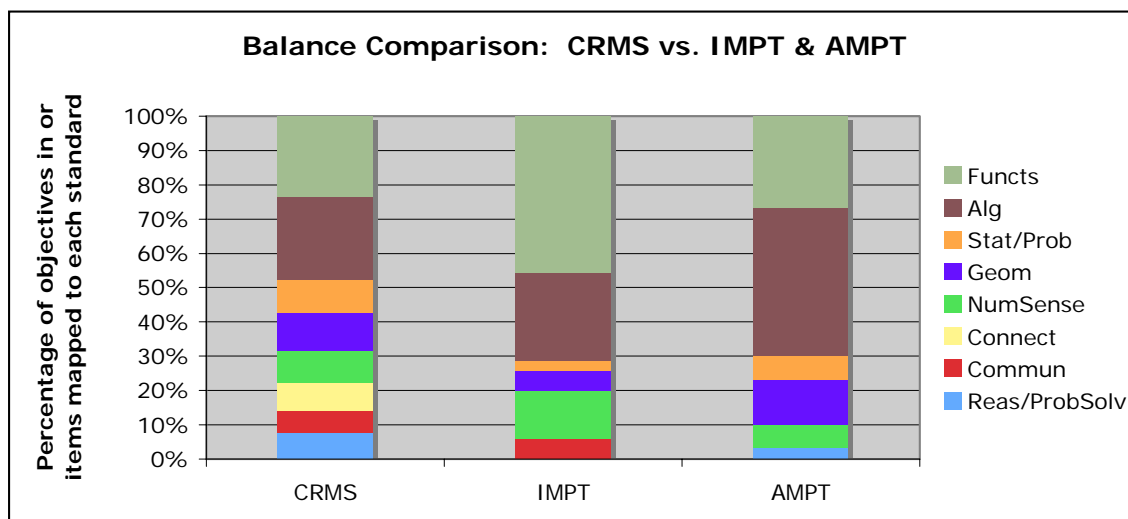
- *Represent basic functions (linear, quadratic, exponential, and reciprocal) and piecewise-defined functions (varying over sub-intervals of the domain) using and translating among words, tables, graphs, and symbols...(extra expectations) and the following advanced functions (cubic, quartic, logarithmic, square root, cube root, absolute value, and rational functions of the type $f(x)=1/x-a$)...*

Although part of Component 8.2 is an *extra expectation*, these two items both deal with the representation of linear functions and therefore align with an aspect of Component 8.2 intended for all students. One of the items on this test was identified as having an inappropriate Source of Challenge other than the mathematics targeted in the question. Item #14 has two algebraically equivalent correct answers among the choices presented to students. That item was mapped to Learning Indicator 5.3.b and is included in the analysis of Balance and Range; it is not included in the review with respect to Level of Challenge, Content Centrality, Performance Centrality, or Level of Demand.

The AMPT has 30 items, all of which demonstrate an acceptable Source of Challenge. Five items on this test had to be mapped to the component level since there was no specific learning indicator to which they aligned. Components 1.3, 5.4, 6.2 and 7.2 were the components to which items were mapped on the AMPT. Eight items on this test were mapped to *extra expectations*. This is entirely appropriate because this test is used to identify possible placement directly into a calculus course.

Balance

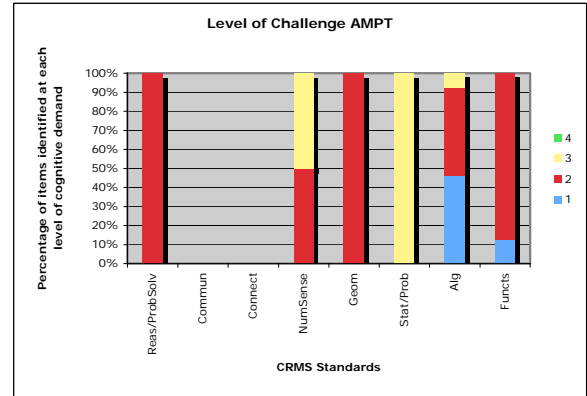
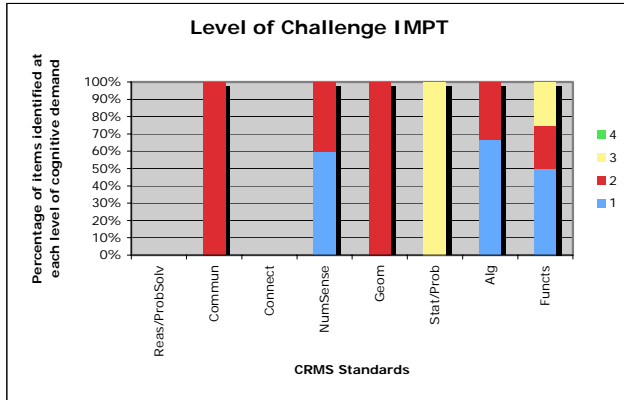
The graph below compares the distribution of the items on each test to the distribution of the components and learning indicators in the CRMS. It should be noted that when percentages are given for a test with only 30 or 35 items, a change in one or two items results in a change in the graph that may appear quite large.



Both the IMPT and AMPT exhibit reasonably good balance across the content standards considering that they have only 35 and 30 items respectively. The Intermediate test has an abundance of items mapped to the Functions standard while the AMPT emphasizes Algebra. Neither test does a particularly good job of balance with respect to the process standards. The IMPT has two items mapped to Communication while the AMPT has one item mapped to Reasoning and Problem-Solving. Different process standards are covered on the two tests; however, students likely take only one of these two placement exams and so will encounter items that address only a single mathematical process area.

Level of Challenge

It is important to recognize the limitations of using percent to report the Level of Challenge. All items on both tests—one item on the IMPT and two items from the AMPT—that were mapped to Statistics and Probability received a Level 3 rating for Level of Demand, which makes this the most challenging standard, though one with obviously limited impact. Similarly, the three items—two on the IMPT and one on the AMPT—that mapped to the process standards were all rated at Level 2 for cognitive demand. More surprising is that this was also the case for the five out of six Geometry items across both tests that could be rated for Level of Demand; one Geometry item on the IMPT was not able to be considered for the analysis of Level of Challenge because it exhibited a Source of Challenge issue (see Source of Challenge section below). The existence of only one Geometry item on the IMPT that could be rated would obviously mean that there could be neither good range of coverage nor range of levels of demand; however, on the AMPT greater variety of cognitive demand would have been both possible and preferable. The remaining three content standards demonstrate a range of levels of demand. The IMPT was judged to be slightly more rote and procedural than the AMPT by virtue of having a higher percentage of its items rated Level 1; this would be expected given the slightly different purposes of the two tests.



Range

Range is a measure of coverage and is measured by the fraction of total objectives (components and learning indicators) within a given standard that are assessed by at least one item. The table below illustrates the range of coverage for each CRMS standard on the IMPT and AMPT. The number of items that mapped to expectations designated as “extra expectations” (those needed by students taking higher level mathematics courses when they enter college) are indicated in italics.

CRMS Standard	Portion of Standards Assessed by the IMPT		Portion of Standards Assessed by the AMPT	
Reasoning/Problem-Solving	0/13	(0%)	1/13	(8%)
Communication	1/11	(9%)	0/11	(0%)
Connections	0/14	(0%)	0/14	(0%)
Number Sense	2/16	(13%)	1/16	(6%)
Geometry	2/19	(11%)	3/15	(20%)
Statistics and Probability	1/16	(6%)	2/16	(13%)
Algebra	8/42	(19%)	13/42 (<i>5</i>)	(31%)
Functions	8/40	(20%)	7/40 (<i>3</i>)	(18%)

Generally, when tests are constructed to align with standards, range values at or above 67% are considered good, while those from 50% to 66% are considered acceptable. Neither of these tests comes close to an acceptable level of coverage. It is difficult to have a good range of coverage when only 30 or 35 items are available to cover 171 objectives. However, coverage could be improved if multiple items were not used to test one learning indicator. This was the case for Learning Indicators 2.1.c, 4.2.a, 7.1.f, 8.1.c, 8.2.a, 8.2.b, 8.4.b, 8.4.c and Component 8.2 on the IMPT. Standard #8 is the Functions standard to which 13 of the 30 items on this test were mapped. Because so many Functions learning indicators had multiple items mapped to them that did not address different aspects of the objective, it would be possible to develop items to assess other aspects of the standards in future versions of the test to improve coverage across all the CRMS standards. For the most part, the AMPT spreads its 35 problems across the standards as well as possible. At most, two items were mapped to the same learning

indicator. In the three cases where that occurred, the two problems addressed different aspects of the learning indicator.

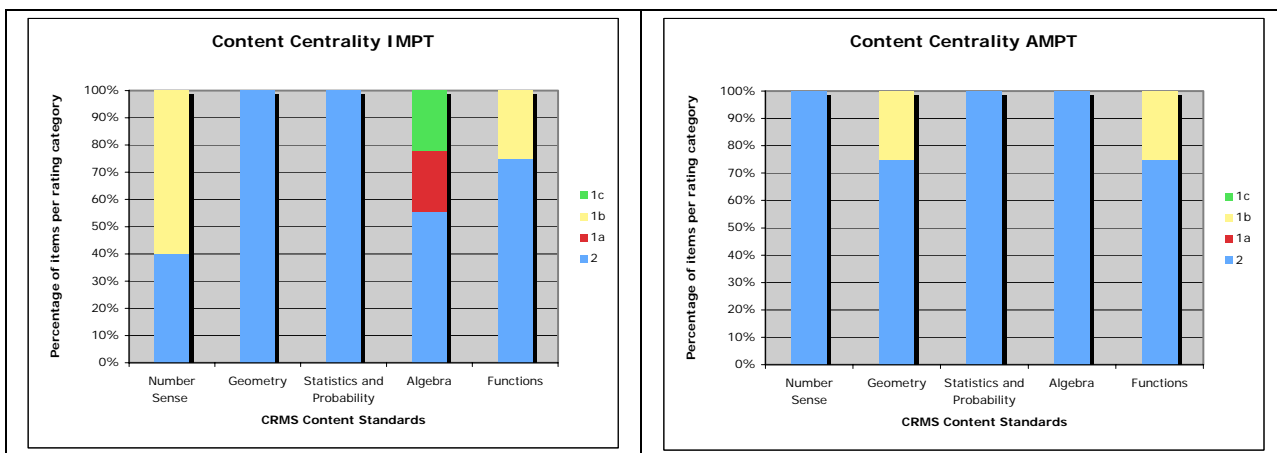
The IMPT addresses only objectives that the CRMS deems core for all students. Given that the AMPT is used to place students into calculus in some cases, it is appropriate that this test assesses some of the *extra expectations* included in the CRMS. It is also appropriate that these would be found in the Algebra and Functions standards, which is the content most needed for success in calculus.

Source of Challenge

All items on the AMPT were deemed to address the major mathematical ideas targeted by the items, and therefore received a rating of *1* for Source of Challenge. One item on the IMPT (#14 mapped to Learning Indicator 5.3.b) had two answer choices that were both right; one was the algebraic equivalent of the other. Each form of the answer might arise depending on the approach the student took to its solution. This item received a Source of Challenge rating of *0*. Any item deemed to have an inappropriate Source of Challenge is considered for analysis with respect to Balance and Range but is not considered for individual item analysis—Content Centrality, Performance Centrality and Level of Demand. This also implies that such an item cannot be used when determining Level of Challenge since this measure, while somewhat global in nature, is based on the distribution of Level of Demand ratings.

Content Centrality

The four-year college and university tests exhibited strong ratings for Content Centrality. Twenty-four of the 34 items (71%) on the IMPT that could be rated in this area were found to measure the most important content that students are expected to know as expressed in the targeted objective; ratings of *2* indicate consistent alignment. Ratings of *1b* indicate that an item assesses only part—and the less central part—of a compound objective, and a rating of *1c* indicates that the objective is so specific that the mapping is only partially consistent. Ratings of *1a* indicate that the objective is too general to be assured of an item's strong alignment. Reviewers assigned a rating of *2* for Content Centrality on the AMPT 87% of the time.



The four items on the AMPT that were not rated 2 were given a rating of *1b*, indicating that they measure part—but not the central part—of a compound objective. The CRMS has many components and learning indicators that are stated in a compound manner so a rating of *1b* for content is almost inevitable for some items. Seven items on the IMPT received a *1b* Content Centrality rating. Unfortunately, some of the items mapping to compound components or indicators on this test did not address different aspects of the objective thus losing an opportunity for stronger content coverage.

The two items receiving a Content Centrality rating of *1a* on the IMPT were both mapped to the same learning indicator, Learning Indicator 7.1.f. As was noted in the Analysis Protocol, all items mapped to this learning indicator received this rating because of the very general nature of the objective. The remaining two test items, both found on the IMPT, received a content rating of *1c*. Item #11 deals with the product of rational expressions where both numerator and denominator are monomials, which is a topic that fits best with Learning Indicator 7.2.e but does not quite match the specifics of that learning indicator.

- Learning Indicator 7.2.e: *Add, subtract, multiply, and divide two rational expressions of the form $a/bx+c$, where a , b , and c are real numbers such that $bx+c \neq 0$ (extra expectations) and of the form $p(x)/q(x)$, where $p(x)$ and $q(x)$ are polynomials.*

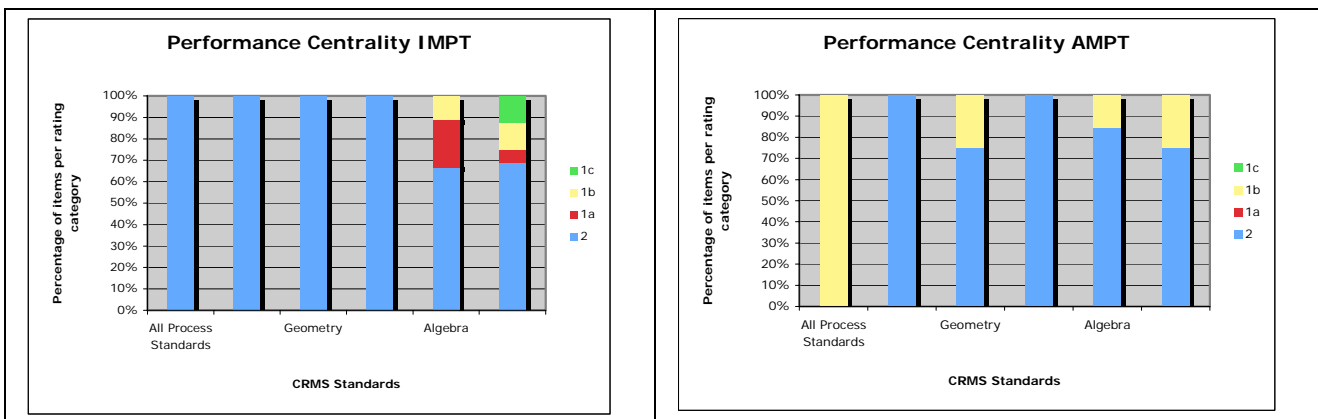
Item #13 presents students with a cubic that would first be factored by removing the common factor before the remaining quadratic factor could be factored into two linear binomial polynomials. The problem asked for one of the two linear binomial factors. While Learning Indicator 7.2.b addressed factoring out the greatest common factor from a polynomial, the best mapping for this item was deemed to be Learning Indicator 7.2.c.

- Learning Indicator 7.2.c: *Factor quadratic polynomials with integer coefficients into a product of linear terms.*

The item does not exactly match this objective and therefore was given a *Ic* Content Centrality rating, which is a comment on the specificity of the objective to which it was mapped.

Performance Centrality

Performance Centrality ratings are strong for both the IMPT and AMPT. As was the case for Content Centrality, the AMPT fared slightly better, with 80% (24 of its 30 items) receiving a rating of 2 in this category. The AMPT’s few remaining items received a Performance Centrality rating of *Ib*. Thirty-four of the IMPT’s 35 items were able to be rated for Performance Centrality. Twenty-six of these items (76%) were rated 2 for Performance Centrality. As was the case with Content Centrality, ratings of 2 indicate consistent alignment, ratings of *Ib* indicate that the item assesses only part—and the less cognitively demanding part—of a compound objective, and a rating of *Ic* indicates that the objective is so specific that the mapping is only partially consistent. A rating of *Ia* signifies that the objective in the CRMS to which an item maps is too general to be assured of the item’s strong alignment.



Reviewers found that both items on the IMPT that map to a process standard require the type of performance described in the targeted objective, resulting in a rating of 2 for Performance Centrality. The only item on the AMPT mapping to a process standard mapped to Component 1.3, a very general objective both in terms of content and performance. As was pointed out in the Analysis Protocol, all items mapped to that component received ratings of *Ia* for both Content Centrality and Performance Centrality. Similarly, there were three items on the IMPT that mapped to learning indicators that reviewers agreed were so general that it was difficult to discern whether the items were clearly consistent with performances required by the objectives. Two of these items (#3 and #9) mapped to Learning Indicator 7.1.f, which was addressed in the Analysis Protocol, and one item (#5) was mapped to Learning Indicator 8.4.c. With respect to this latter learning indicator, reviewers agreed that abstracting models and interpreting solutions are actions performed in so many mathematics problems that it is difficult to identify exactly what is intended by this objective.

- Learning Indicator 8.4.c: *Abstract mathematical models from word problems and interpret solutions in the context of these source problems.*

The two items from the IMPT that were given a *1c* performance rating were mapped to two learning indicators in the Functions standard, which were both cited in the Analysis Protocol:

- Learning Indicator 8.1.c: *Understand and interpret function notation, particularly as it relate to graphic displays of data.*
- Learning Indicator 8.2.a: *Evaluate functions to generate a graph.*

While the two items appropriately mapped to their respective learning indicators, neither of them involved a graph.

Level of Demand

Across both tests, the Level of Demand was varied although one might have expected more Level 3 items on the AMPT than was the case. Level 3 items are the most cognitively demanding of the three types of items generally found on multiple-choice tests, requiring that students use reasoning and planning skills and employ a higher level of thinking than required by the other two levels of items typically found on such tests. Level 2 items require students to make some decisions as to how to approach a problem, whereas Level 1 items require students to recall information or perform a well-known algorithm or procedure.

Because one item exhibited an inappropriate Source of Challenge, only 34 of the 35 items on the IMPT could be rated for Level of Demand. Of those, 50% percent received a Level 1 rating in this area, just over 35% received a Level 2, and the remaining five items were deemed to require the type of reasoning exemplified by a Level 3 rating. All 30 of the AMPT items were able to be rated for Level of Demand. Just over 23% received a Level 1, just over 63% were rated Level 2, and the remaining four items, one of which was mapped to *extra expectations*, were rated Level 3 for Level of Demand.

The range of levels on these two tests indicates that students will encounter content and performance demands that require decision-making and strategizing, in addition to recall of information and automaticity. This is a positive feature for a test to exhibit and indicates that the IMPT and AMPT reflect the breadth and richness of thinking suggested by the CRMS and needed for future success in postsecondary mathematics.

APPENDIX C: WHATCOM COMMUNITY COLLEGE

Whatcom Community College submitted four placement tests for alignment. Entering students are given the Math Placement Survey. On the basis of the survey, they will be given the opportunity to take one of the four placement tests. The Basic Mathematics test is used to determine placement into courses below the college's Algebra I course. The Beginning Algebra test is used with students who, on the basis of the survey, are targeted for placement into the Math 97-98 sequence—Elementary Algebra I and Elementary Algebra II. Students targeted for the Math 98-99 sequence—Elementary Algebra II and Intermediate Algebra—are tested with the Intermediate Algebra placement exam. The last half of the Beginning Algebra test is identical with the first half of the Intermediate Algebra Test. This seems appropriate since the potential placement sequences have one course in common. Finally, for students deemed ready to take “College Level” courses—Survey of Math, Precalculus I or II, Algebra Applications for Economics and Business, or Calculus—the placement test used is the Pre-Calculus test.

Map of Test Items to the College Readiness Mathematics Standards

The Basic Mathematics test is comprised of 40 items. Three of these are true-false items; the other 37 are presented in multiple-choice format. Of those 37, ten offer three potential answer choices while the rest offer four. There were three items for which no component or learning indicator could be found that offered an appropriate match:

- Item #11 asks students to write a three-digit number in expanded form.
- Item #18 requires labeling a given number as prime or not prime.
- Item #20 requires the determination of the least common multiple of a set of three small numbers.

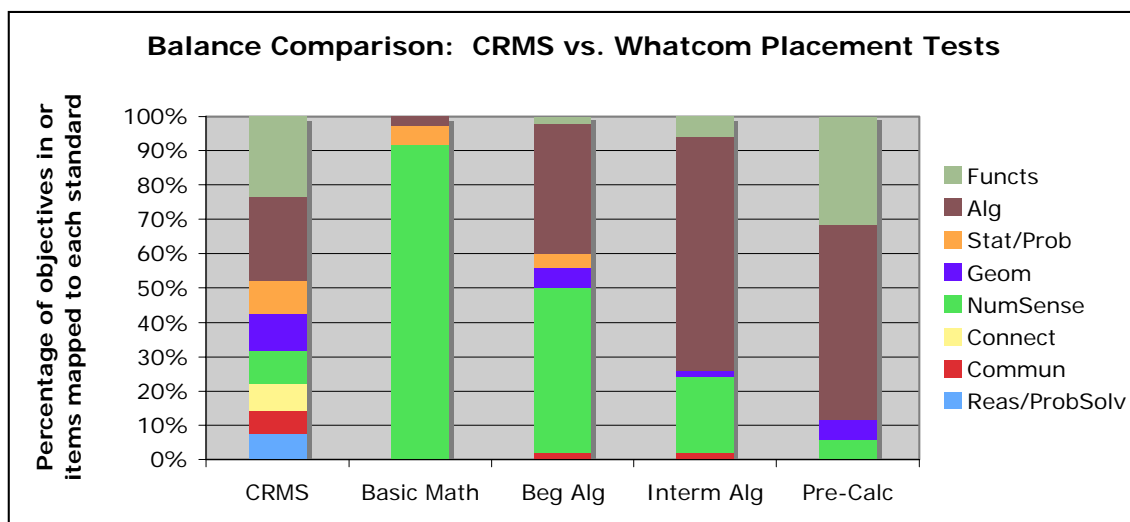
These items are not used in the alignment analysis, thus leaving 37 items that could contribute to the determination of the balance and range of the test. Two of the 37 items were judged to have an inappropriate Source of Challenge. These items were not included in the analysis of Content Centrality, Performance Centrality, or Level of Demand. Since the Level of Challenge of each standard relies solely on the ratings for Level of Demand, the analysis for this section is also based on only 35 items.

Beginning Algebra and Intermediate Algebra each contains 50 multiple-choice questions with five possible answers. Items #26 – 50 on Beginning Algebra are identical to questions #1 – 25 respectively on the Intermediate Algebra test. All items could be mapped to components or learning indicators of the CRMS, and all were used in the full alignment analysis although two items unique to the Beginning Algebra test were somewhat problematic in the eyes of the reviewers. Because it is unlikely that a particular student would take both the Beginning Algebra and Intermediate Algebra tests, both unique and common items were included when reporting results for each test.

The Pre-Calculus placement exam consists of 37 multiple-choice questions, each of which has five answers from which to select. All items were mapped to the CRMS and were included in the analysis.

Balance

Across all four tests, only one item—#50 on Beginning Algebra, which is the same as #25 on Intermediate Algebra—mapped to any of the CRMS process standards. While the Basic Mathematics test is the most uniform, none of the four tests exhibits the breadth and depth of mathematical topics evident in the CRMS. The graph below illustrates a shift in emphasis from one Whatcom test to the next. Across the tests, Number Sense becomes less of a focus while the Algebra and Functions standards receive increasing emphasis.



Balance within each standard is also an area for concern with these tests. The Basic Mathematics test has 34 of its 37 items mapped to the Number Sense standard. Of these, 24 are mapped to Component 4.2 since the computation they call for involves only a single operation with whole numbers or simple fractions. While all four basic operations are tested, the items on the Basic Mathematics test miss the opportunity to assess computation with variety of number types implied by the objective.

- Component 4.2: *Accurately and efficiently compute with real number in all forms, including rational exponents and scientific notation.*

Although not as pronounced, the Beginning Algebra test is similarly unbalanced across the objectives in the Number Sense standard. Six of its 24 Number Sense items mapped to Component 4.2, and 10 mapped to Learning Indicator 4.2.a. Those items mapped to Learning Indicator 4.2.a differ only in that they involve at least two arithmetic operations, but they still only address the most basic types of numbers. Even on the Intermediate Algebra test, nine of the 11 Number Sense items (82%) are mapped to Learning Indicator 4.2.a while only one (#36) involves even an irrational number.

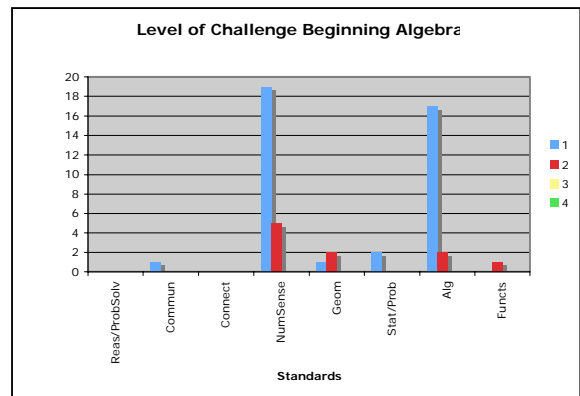
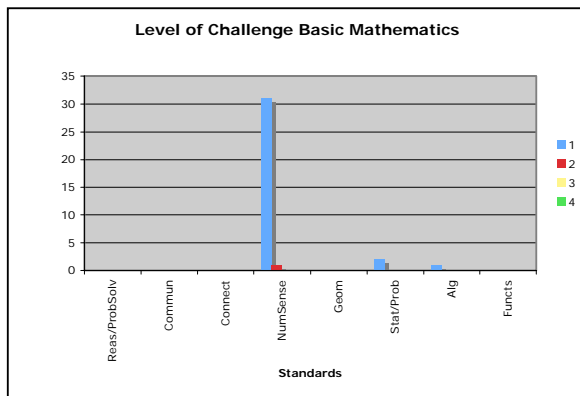
Balance across the Algebra standard fares somewhat better in the Beginning Algebra and Intermediate Algebra tests. While multiple items are mapped to Learning Indicator 7.2.a on both tests, the items address different aspects of that standard.

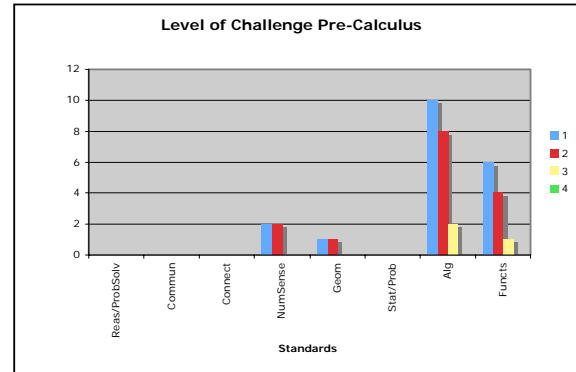
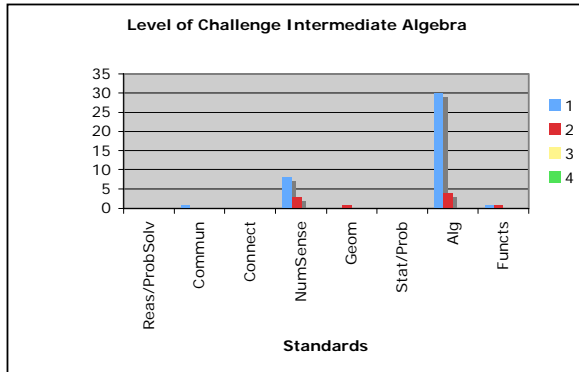
- Learning Indicator 7.2.a: *Find the sum, difference, or product of two polynomials, then simplify the result.*

The balance within standards is not a problem for the Whatcom Pre-Calculus test.

Level of Challenge

On the Basic Mathematics test, only 35 of the 37 mapped items could be used to analyze the Level of Challenge. Across all standards, this test included only one item—item #14 mapped to Component 4.3—that rose above the level of rote application of knowledge or procedures. Fully 40 out of 50 items (80%) on the Beginning Algebra test also required only this level of cognitive demand. The remaining items were, however, spread across four content standards—Number Sense, Geometry, Algebra, and Functions—offering a minimal range of challenge within each of these standards. No items on either the Basic Algebra or the Intermediate Algebra tests require students to make decisions or strategize—actions that would have caused the item to be given a Level of Demand rating of Level 3. Eighty percent of the Intermediate Algebra items received a Level 1 rating, which implies the lowest level of cognitive demand. The other 20% received a Level 2 rating. The Pre-Calculus test had two items (#9 and #28) that mapped to the Algebra standard and one item (#6) that mapped to the Functions standard that received a Level 3 rating for Level of Demand. All three of these items address learning indicators or portions of learning indicators that are identified as *extra expectations* within the CRMS (Learning Indicator 7.3.h, 7.3.k, and 8.2.c respectively). While not high in number, these three items do provide a range of challenge within the Algebra and Functions standards and demonstrate that the Pre-Calculus test has the best Level of Challenge across the four standards represented by items on that test. However, because only four standards are represented, it is difficult to be completely satisfied with the Level of Challenge across the standards.





Range

Range is a measure of coverage and is measured by the fraction of total objectives (components and learning indicators) within a given standard that are assessed by at least one item. The table below illustrates the range of coverage for each standard on each of the two Whatcom tests. The number of items that mapped to expectations designated as *extra expectations*—those needed by students taking higher-level mathematics courses when they enter college—is indicated in italics.

CRMS Standard	Portion of Standards Assessed by Basic Mathematics		Portion of Standards Assessed by Beginning Algebra		Portion of Standards Assessed by Intermediate Algebra		Portion of Standards Assessed by Pre-Calculus	
Reasoning/Problem-Solving	0/13	(0%)	0/13	(0%)	0/13	(0%)	0/13	(0%)
Communication	0/11	(0%)	1/11	(9%)	1/11	(9%)	0/11	(0%)
Connections	0/14	(0%)	0/14	(0%)	0/14	(0%)	0/14	(0%)
Number Sense	5/16	(31%)	7/16	(44%)	3/16	(19%)	2/16 (2)	(13%)
Geometry	0/19	(0%)	1/19	(5%)	1/19	(5%)	1/15 (1)	(1%)
Statistics and Probability	1/16	(6%)	2/16	(13%)	0/16	(0%)	0/16	(0%)
Algebra	1/42	(2%)	9/42	(21%)	13/42	(31%)	12/42 (7)	(28%)
Functions	0/40	(0%)	1/40	(3%)	2/40	(5%)	7/40 (4)	(18%)

Generally, when tests are constructed to align with standards, range values at or above 67% are considered good, while those from 50% to 66% are considered acceptable. None of these tests comes close to an acceptable level of coverage for any standard. The lack of balance within a standard on some of the tests that was noted above contributes to the low range of coverage within those standards. Because only two items mapped to the Geometry standard on the Pre-Calculus test, coverage of the 19 objectives in that standard is necessarily low. However, the imbalance of items also offers the potential for a good range of coverage of those standards that are over-represented by items on the test. Unfortunately, that potential is not realized. For example, the 92% (34 out of the 37 mapped items) on the Basic Mathematics test mapped to Number Sense provides an excellent opportunity for full coverage of the 16 objectives in that standard. The fact that 24 of the items addressed Component 4.2 and an additional five addressed Learning

Indicator 4.2.a means that coverage of the remaining 14 objectives in that standard is unnecessarily light.

Source of Challenge

Occasionally an item contains some aspect that may provide a challenge to students that comes from something other than the targeted mathematics content; such an item is rated 0 with respect to Source of Challenge. The only placement test from Whatcom Community College that received a 0 in this category was the Basic Mathematics test where two such items were identified.

- Item #1, mapped to Component 4.2, asks students to determine the monthly payment for a truck with a given cost. Reviewers felt that there was the potential for students to be confused about whether interest should be included since such loans are rarely interest-free and might select a higher payment to account for the interest.
- Item #10, mapped to Learning Indicator 4.2.a, gives information about the percent of people visiting several venues but asks “How many people...” went to a particular venue. The answers could only be given in terms of a percent.

Items identified with a Source of Challenge issue are not included in the alignment analysis at the item-specific level; that is, they are not considered for Content Centrality, Performance Centrality, or Level of Demand. As was mentioned above, they are necessarily ignored as well for the purpose of determining the Level of Challenge of the standards. In the case of the Basic Mathematics test, exclusion of these two items made no effective difference to the overall messages of the alignment analysis for the test.

Content Centrality

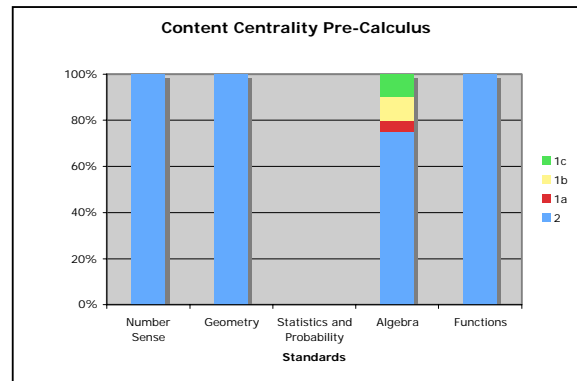
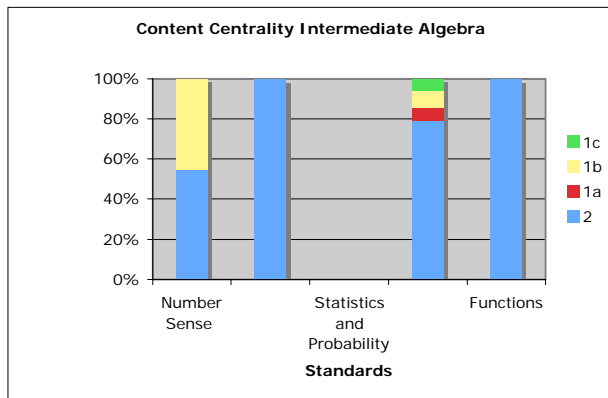
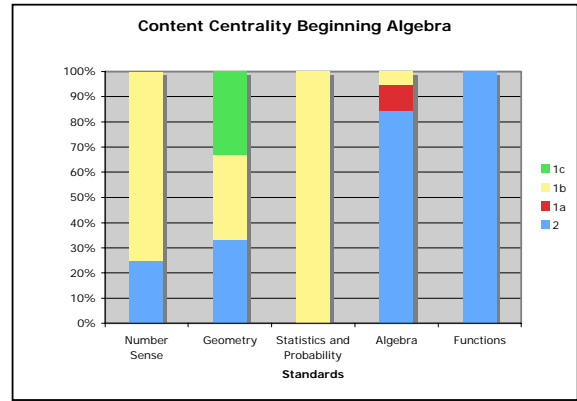
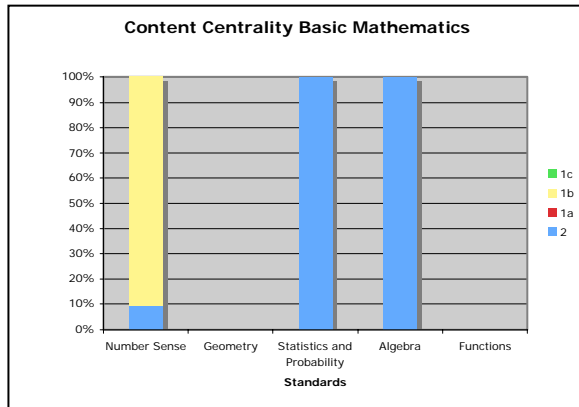
The review team assigned a Content Centrality rating to the single item mapped to a Process Standard—#50 on the Beginning Algebra test, which is identical to #25 on the Intermediate Algebra test—that mapped to Learning Indicator 2.1.b in the process standards even though it did not strictly speaking address mathematical content. This item received a rating of *Ib*, which indicates that it addresses only part—and the less central part—of the objective.

- Learning Indicator 2.1.b: *Recognize and explain the meaning of information presented using mathematical notation.*

Because the item only requires students to identify the ordered pair that represents the coordinates of a given point on a coordinate plane, reviewers agreed that the item aligned only to a very small and very easy part of the objective.

As can be seen in the graphs that follow, the strength of content alignment varies by test and by content domain. Ratings of 2 indicate consistent alignment; ratings of *Ia* indicate that the objective is too general to be assured of an item’s strong alignment; ratings of *Ib*

indicate that the item assesses only part—and the less central part—of a compound objective; and ratings of 1c indicate that the objective is so specific that the mapping is only partially consistent.



The two more advanced Whatcom tests exhibited relatively strong ratings for Content Centrality. Seventy-four percent of the items on the Intermediate Algebra test received a rating of 2. Eight items mapped to the content standards received a rating of 1b for Content Centrality. Five of the eight items (#16, #18, #30, #34, and #46) mapped to Learning Indicator 4.2.a are rated 1b because they address computation that involves only whole numbers or simple fractions. The remaining three items (#8, #32, and #42), mapped to Learning Indicator 7.2.f, involve expressions with only integer—rather than rational—exponents.

- Learning Indicator 7.2.f: *Simplify products and quotients of single-term expressions with rational exponents (rationalizing denominators not necessary).*

The two items on this test that received a Content Centrality rating of 1c (#2 and #3) both mapped to Learning Indicator 7.2.a. They received a 1c rating because they involve operations combining more than two polynomials while the learning indicator limits the operations in this way.

- Learning Indicator 7.2.a: *Find the sum, difference, or product of two polynomials then simplify the result.*

Both items receiving a rating of *1a* (#28 and #31) were mapped to Learning Indicator 7.1.f, cited in the Analysis Protocol as being so general that all items mapped to it received a *1a* rating for both content and performance centrality.

On the Pre-Calculus test, 32 of the 37 items (86%) were deemed to be consistently aligned to the objectives to which they were mapped. The one item (#33) mapped to Learning Indicator 7.1.f received a rating of *1a*. The two items (#31 and #34) receiving a rating of *1c* both involved systems of equations with three variables while the learning indicator to which they were mapped—Learning Indicator 7.3.c—specifically refers to systems involving two variables. The remaining two items (#14 and #19) received a Content Centrality rating of *1b*. No learning indicator matched the content of these items because they involved finding solutions of cubic equations; therefore, they were mapped to Component 7.3, which encompasses so many types of equations that it was difficult to say that these items addressed the central part of the objective.

- Component 7.3: *Solve various types of equations and inequalities numerically, graphically, and algebraically; interpret solutions algebraically and in the context of the problem; distinguish between exact and approximate answers.*

Only six of the 35 items (17%) on Whatcom’s Basic Mathematics that could be given a rating for Content Centrality received a rating of 2. As has been noted elsewhere, this low content alignment is mainly due to the type of numbers involved in the very basic computation problems included in the test. While the items mapped to standards other than Number Sense receive strong ratings for content centrality, there are only three such items, thus minimizing the effect on content alignment across the test.

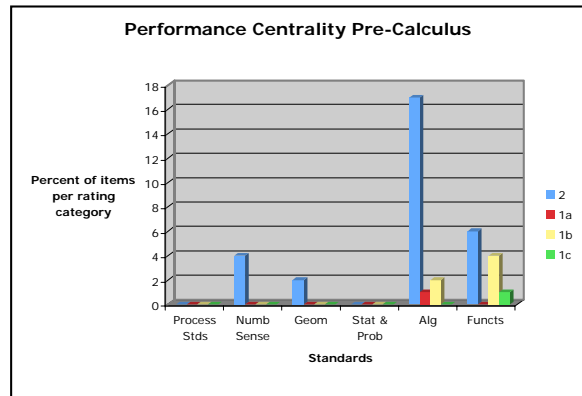
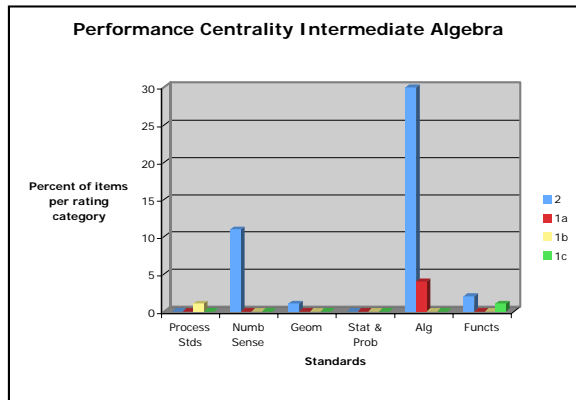
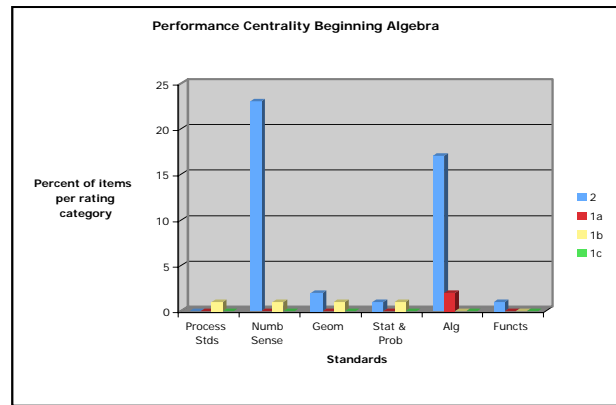
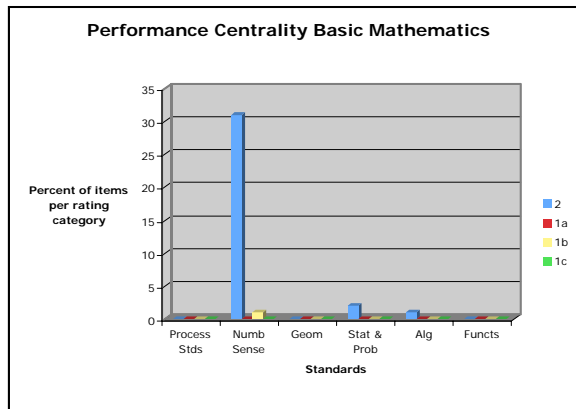
Forty-eight percent of the items on the Beginning Algebra test were rated 2 for content centrality, indicating relatively poor content alignment to the CRMS. The type of numbers involved in the various computations can explain a rating of *1b* for 18 of the 24 items mapped to Number Sense. Item #14 received a rating of *1b* and item #23 received a rating of *1c*—both of these items mapped to Learning Indicator 5.3.d. Item #14 addressed only the easiest part of the content covered by the objective—addressing the perimeter of a rectangle—while item #23 went beyond the specific limits of the objective—involving a figure that is a composition of rectangles.

- Learning Indicator 5.3.d: *Calculate the area and perimeter of circles, triangles, quadrilaterals, and regular polygons.*

The two items that mapped to the Statistics and Probability standard each aligned to the easiest part of the compound content indicated in the objective to which they were mapped. The two Algebra items both mapped to Learning Indicator 7.1.f and so received a *1a* rating.

Performance Centrality

As was the case with Content Centrality, ratings of 2 indicate consistent alignment, ratings of *1b* indicate that the item assesses only part—and the less cognitively demanding part—of a compound objective, and a rating of *1c* indicates that the objective is so specific that the mapping is only partially consistent. A rating of *1a* signifies that the objective in the CRMS to which an item maps is too general to be assured of the item’s strong alignment. It is evident from the graphs below that all four tests demonstrated strong alignment with the CRMS with respect to Performance Centrality.



Ninety-one percent of the items that could be given a performance rating on the Basic Mathematics test received a 2. Items on each of the Beginning Algebra and Intermediate Algebra tests received ratings of 2 in this area 88% of the time. The Pre-Calculus test had the lowest performance alignment of the four tests, but even on this test, 29 of 37 items (78%) received a Performance Centrality rating of 2. The only items receiving a rating of *1a* on any of the tests were mapped to the very generally stated Learning Indicator 7.1.f. Both items receiving a rating of *1c*—one on the Intermediate Algebra test and one on the Pre-Calculus test—involved evaluation of a function but not for the purpose of generating a graph, which was the limitation placed on the operation by Learning Indicator 8.2.a (*Evaluate functions to generate a graph*).

Level of Demand

Perhaps the area for greatest concern is the level of cognitive demand required of students taking any of the Whatcom tests. Achieve’s rating scale defines four levels of cognitive

demand although a multiple-choice format generally only allows for three to be assigned. Level 3 items are the most cognitively demanding of the three types of items, requiring that students use reasoning and planning skills and employ a higher level of thinking than the other two levels. Level 2 items require students to make some decisions as to how to approach a problem, whereas Level 1 items require students to recall information or perform a well-known algorithm or procedure. The Whatcom tests assess a low level of cognitive demand for instruments that are used to guide the course placement of students entering college.

Of the 34 of 35 items on the Basic Mathematics test that were mapped and did not exhibit a Source of Challenge issue, fully 97% received a Level 1 rating for Level of Demand. The one remaining item received a Level 2 rating and in that case only minimally because the problem was presented in a context that would require students to sort out where to put the numbers before performing a simple computation. Eighty percent of the items on both the Beginning Algebra test and the Intermediate Algebra test received a Level 1 rating while the remaining items rose only to Level 2. Only on the Pre-Calculus test were any items rated Level 3, and even then, only three items received that rating. Furthermore, more than half (19 out of 37 items) were rated Level 1 for Level of Demand on a test that is intended to indicate placement into upper-level mathematics courses.

On all of these tests, the cognitive demand on students is too often at a recall or rote level. These tests do not provide information that would indicate the level of thinking generally required in college-level mathematics courses.

APPENDIX D: BIOGRAPHIES

ACHIEVE STAFF

KAYE FORGIONE, SENIOR ASSOCIATE, MATHEMATICS, ACHIEVE

Kaye Forgione joined Achieve as senior associate for mathematics in March 2001 where she leads Achieve's Standards and Benchmarking Initiatives involving mathematics. Prior to joining Achieve, Kaye served as assistant director of the Systemic Research Collaborative for Mathematics, Science and Technology Education (SYRCE), a project at the University of Texas at Austin funded by the National Science Foundation. Her responsibilities at the University of Texas also included management and design responsibilities for UTeach, a collaborative project of the College of Education and the College of Natural Sciences to train and support the next generation of mathematics and science teachers in Texas. Before her work at the University of Texas, Kaye was director of academic standards programs at the Council for Basic Education, a nonprofit education organization located in Washington, DC. Prior to joining the Council for Basic Education in 1997, Kaye worked in the K–12 arena in a variety of roles, including several leadership positions with the Delaware Department of Education. Kaye began her education career as a high school mathematics teacher. She taught mathematics at the secondary and college levels as part of adult continuing education programs. Kaye received a bachelor's degree in mathematics and education from the University of Delaware, a master's degree in systems management from the University of Southern California, and a doctorate in educational leadership from the University of Delaware.

LAURA MCGIFFERT SLOVER, DIRECTOR, CONTENT & POLICY RESEARCH, ACHIEVE

Laura McGiffert Slover is director of Content & Policy Research at Achieve, where she has senior responsibility for overseeing a number of Achieve's major initiatives. She supervises Achieve's Benchmarking Initiative, leads its work with states on building mathematics capacity, and oversees the organization's research agenda. Laura has extensive experience reviewing academic standards and education policies in the United States and abroad, and she has written a number of reports and articles on the topic. Before joining Achieve in 1998, Laura was a high school English teacher in Eagle County, Colorado, where she was involved in the district's early efforts to develop standards and benchmark assessments. She also taught writing and composition at Colorado Mountain College. A native Washingtonian, Laura earned a bachelor's degree in English and American Literature from Harvard University; a master's in Education Curriculum and Instruction from the University of Colorado at Boulder; and a master's in Education Policy from Georgetown University. She is a mentor and a member of the Board of Directors of Project Northstar, an organization that provides mentoring and tutoring to homeless and at-risk students in the District of Columbia.

CONTENT EXPERTS

SUSAN K. EDDINS

Susan K. Eddins taught students in kindergarten through college for over 30 years. She is the recipient of several honors for her teaching, including the Presidential Award for Excellence in Mathematics Teaching, and she is a National Board Certified Teacher in Adolescent and Young Adult Mathematics. Ms. Eddins is now retired having been a faculty member, an Instructional Facilitator, and the Curriculum and Assessment Leader in mathematics at the Illinois Mathematics and Science Academy, where she taught since the school's inception in 1986. She has served in leadership capacities in several professional organizations, notably as a member of the Board of Directors of the National Council of Teachers of Mathematics (NCTM). Ms. Eddins was a member of the 9–12 writing group for NCTM's *Principles and Standards for School Mathematics*. She is co-author of a chapter in NCTM's *Windows of Opportunity* and is a co-author of *UCSMP Algebra*. She is a past panel member and editor of NCTM's *Student Math Notes* and has authored several articles in refereed journals. More recently, in addition to numerous workshops and presentations, her most extensive work has been in the area of standards development, standards review, and alignment of standards to assessments. For Achieve, she has reviewed academic standards or assessments from Alaska, Illinois, Indiana, Minnesota, New Jersey, Oregon, Pennsylvania, Texas and Washington. Ms. Eddins holds bachelor's and master's degrees in mathematics.

MARY LYNN RAITH

Mary Lynn Raith received her B.S in mathematics from Indiana University at Pittsburgh and her M.Ed. in mathematics education from the University of Pittsburgh. She is recently retired from the position of Mathematics Specialist in the Division of Instructional Support of the Pittsburgh Public Schools. As such, her responsibilities included leadership roles in curriculum development, textbook selection, design of alternative assessments, in-service program design and implementation, and coordination of mathematics programs across levels and schools. Ms. Raith was also the Co-Director of the Pittsburgh Reform in Mathematics Education project (PRIME), a K-12 professional development system. She has also been involved with a number of national projects, including the development of both the New Standards Reference Examination and the Portfolio project for the middle grades, the Assessment Communities of Teachers project (ACT), and the Alternative Assessment in Mathematics project (A²IM). She has also worked extensively with both NCTM and NCEE on its America's Choice school design and has presented at numerous national conferences.