OFFICE OF EDUCATIONAL ASSESSMENT
Improving Learning Through Assessment

OEA Report 08-01

General Mathematics Placement Test (MPT-G): Initial Test Development

Debbie McGhee, Jon Peterson, Jerry Gillmore, and Nana Lowell
August 2008

INTRODUCTION

The Mathematics Placement Testing (MPT) program was developed by a collaboration of math faculty at Washington State public universities to assist students, together with their academic advisers, in selecting first-year mathematics courses for which they are best prepared. The MPT program currently offers two tests: the Intermediate test (MPT-I) for placement into precalculus, and the Advanced test (MPT-A) for placement into calculus. The program began somewhat informally in the mid 1980's, but was standardized in 1994 with the development of the current test versions and score database. The MPT is managed by the UW Academic Placement Testing Program (APTP) and each year tests about 10,000 undergraduates admitted to Washington public universities.

Over the past several years, there has been increasing concern at both the state and national level about mathematics preparation among high school students. In Washington State, the Transition Math Project (TMP)[1] has taken a lead role in shaping statewide efforts to improve the linkage between K-12 and post-secondary mathematics education. The TMP initiated a collaborative project in June 2004 to define standards for college readiness in mathematics, and the final College Readiness Mathematics Standards (CRMS)[2] were released in March 2006. As an extension of this work, TMP and the UW Office of Educational Assessment (OEA) proposed that the APTP Math Placement tests be more closely aligned with the new CRMS to provide a means of assessing student readiness for college level mathematics, and that the tests be made available to students at two-year post-secondary institutions and in the high schools. The MPT is uniquely appropriate for college readiness testing because the existing statewide testing infrastructure can easily be expanded to other student groups, and because the MPT-I is already functioning as a de facto measure of college readiness in admitting students to entry level college mathematics courses.

The TMP/OEA proposal served as the basis for the Second Substitute House Bill 1906[3] passed during the 2007 legislative session. HB 1906 stipulated that the MPT program be revised as described above, with the added requirement that the test(s) should yield a single "college ready" cutoff score indicating whether a student is eligible to enroll in credit-bearing college math courses. OEA is carrying out this work with TMP sponsorship and in collaboration with representatives from post-secondary and K-12 sectors. This report describes the initial test development of the General Math Placement Test, or MPT-G, which is intended to be an addition to the MPT battery rather than a replacement of any existing test.

[1] http://www.transitionmathproject.org/
[2] http://www.transitionmathproject.org/standards/docs/crs_march23_2006.pdf
[3] http://apps.leg.wa.gov/documents/billdocs/2007-08/Pdf/Bills/House%20Passed%20Legislature/1906-S2.PL.pdf

INITIAL TEST SPECIFICATIONS

With the assistance of TMP, OEA formed a collaborative College Readiness Mathematics Test (CRMT) working group[4] to create a new Math Placement test for use by post-secondary and high school students. The test would 1) be aligned with the CRMS, 2) be appropriate for use in placing students into general college-level mathematics courses, 3) yield a single "college ready" cutoff, and 4) provide the basis for advisory use with 11th grade students. The working group met twice in late spring and early summer 2007 to determine test structure and to write sample items "operationalizing" the CRMS components.[5]

Although we originally anticipated that the content of the existing MPT-I could be broadened to accommodate placement into general math courses, the CRMT working group rejected this alternative based on concerns that it would reduce the predictive validity of the test relative to precalculus courses due to differences in content emphases. Instead, the group proposed development of a separate General Math Placement test (MPT-G) for entry level college courses outside the precalculus-calculus track. The test would be broader in content than the MPT-I, but would be of the same level of difficulty and thus provide the same indication of college readiness. The MPT program would therefore offer three placement tests (MPT-G, MPT-I, and MPT-A) that students would select based on the course for which they wished to register. Institutions would continue to utilize different placement cutoffs according to varying course characteristics, but would agree on a single "college ready" cutoff for at least the two lower level tests.

Content

To define the content and initial table of specifications for the MPT-G, we asked participants in the working group to rate the importance of each CRMS Content Component in assessing readiness for general, entry-level mathematics courses.[6] We also took advantage of the opportunity to ask the group to provide similar, comparative, ratings for precalculus courses to support renewal of the MPT-I. As noted in OEA's 2006 alignment study,[7] items on the current MPT-I (Version F) were not evenly distributed across CRMS Content Standards, and we wanted to evaluate the appropriateness of the current MPT-I content. Table 1 shows the actual item distribution for the MPT-I (Version F) and proposed item distributions for the new MPT-G and MPT-I based on the ratings provided. Working group participants considered the five CRMS Content Standards to be of approximately equal importance in assessing general college readiness, but gave more emphasis to *Algebra* and *Functions* in assessing readiness for precalculus.

We also asked members of the working group to provide operational definitions of the CRMS Components by writing sample items. At the first working meeting, participants created items for each of the Content Standards, and at the second meeting they focused on items relating to *Number Sense*, *Geometry*, and *Probability & Statistics* because these Standards were underrepresented in the existing MPT item pool. Participants were encouraged to thread the three Process Standards within these newly developed items.

---

[4] The group included representatives from two-year and four-year post-secondary schools as well as K-12.

[5] http://www.washington.edu/oea/services/testing_center/crmt/about_crmt.html

[6] Components referencing "extra expectations" were not included.

[7] http://www.washington.edu/oea/pdfs/reports/OEAReport0606.pdf

Table 1. Item frequency distributions for proposed MPT-G and MPT-I

CRMS Content Standards and Components	Proposed MPT-G	Proposed MPT-I	Current MPT-I
Number Sense	7	8	5
4.1 Understand concept of real numbers	2	2	1
4.2 Compute with real numbers	3	3	4
4.3 Apply estimation strategies	2	3	0
Geometry	6	5	3
5.1 Make and test conjectures about 2-D figures	0	0	0
5.2 Represent physical situation using 2-D figures	1	2	0
5.3 Use properties of figures to draw and justify conclusions	4	3	3
5.4 Apply right triangle relationships to solve problems	1	0	0
Probability & Statistics	6	0	1
6.1 Use probability to solve problems involving uncertainty	2	0	0
6.2 Develop displays to represent and study data	1	0	1
6.3 Develop and evaluate inferences based on data	2	0	0
6.4 Create and evaluate linear models	1	0	0
Algebra	8	10	9
7.1 Recognize and use concepts to simplify expressions	2	3	2
7.2 Combine and simplify algebraic expressions	2	3	2
7.3 Solve equations and inequalities	4	4	5
Functions	8	12	17
8.1 Recognize functional relationships	3	3	4
8.2 Represent basic functions	3	4	5
8.3 Analyze and interpret features of a function	0	3	1
8.4 Model situations and relationships	2	2	7
TOTAL	35	35	35

Note. CRMS Component descriptions are abbreviated.

Format

Two- and four-year institutions employ different administrative mechanisms to deliver placement tests. The majority of students who enter four-year institutions do so in the fall, allowing large group testing sessions in spring and summer and the corresponding efficiencies of paper-pencil testing. In contrast, students enter two-year schools throughout the year, and individual testing is more appropriate. Because of this testing pattern and the availability of commercial testing applications, the majority of the two-year institutions have implemented computer-based testing. To best serve these schools, we set as a long range goal the development of online versions of all Math Placement tests. Initially, however, the MPT-G would be developed in the same format as currently used for the MPT-I and MPT-A, namely as a one-hour, paper-pencil, timed test.

ITEM DEVELOPMENT AND TESTING

Our initial item pool consisted of previously unused (Version H) or infrequently used (Version G) MPT-I items and the sample items created by the CRMT working group. OEA staff categorized each item according to the College Readiness Mathematics Standards (both Process and Content**)** to which it related, and identified the primary Standard for items with multiple mappings. We reviewed and modified items written by the working group, determined where additional items were needed to yield an item pool with the content distribution specified in Table 1, and held a second working group meeting in which participants wrote new items for

underrepresented content areas. In total, we chose or developed 120 new items (24 per Content Standard), approximately twice the number needed for two parallel forms of the MPT-G.

We next carried out an empirical pretest to determine the performance characteristics of each item. The goals of the pretest were to identify a sufficient number of high quality items to form two parallel forms of the MPT-G and to begin populating a large item bank as a basis for ongoing test replacement for the MPT-G and MPT-I.

Pretest Methodology

Instruments

The pretest item pool consisted of the 120 new content (CRMT) items and six benchmark items from the current MPT-I (Version F). Benchmark items were included to allow us to equate the difficulty of new items to those from existing Math Placement tests. Candidate benchmark items were identified as those for which the correlation (item-rest) between the item score (0 or 1) and the score on the remaining test items was greater than or equal to .40, computed from all MPT-I (Version F) tests administered between January 1, 2003 and September 15, 2007 (N = 28,703). Seven items met this criterion (Table 2), and we selected the set of six items that produced the maximum discrimination between the lowest and middle tertile groups (Subset 1 in Table 3).

Table 2. MPT-I (Version F) items with item-rest $r \geq .40$

Item	Item-Rest *r*	Mean
Item 1	.51	.53
Item 22	.49	.55
Item 29	.48	.72
Item 2	.44	.38
Item 4	.43	.54
Item 31	.41	.40
Item 23	.40	.59

Table 3. Comparative statistics for two sets of MPT-I items.

Six item subset	Cronbach's α	Mean number correct	Discrimination between lowest and middle tertile groups
Subset 1: 1,2,4,22,29,31	.685	3.13	M_{diff} = .31, d = .66
Subset 2: 1,2,22,23,29,31	.669	3.18	M_{diff} = .28, d = .59

Utilizing the 120 CRMT items and six benchmark items, we created two 30-item alternate test versions for each of the five Standards. The first version consisted of two six-item blocks of content items, followed by the six benchmark items, and then two additional six-item blocks of content items. The second version contained the same items as the first, but the order of the four content blocks was reversed to control for order effects. The item layout for each test version is shown in Appendix A.

Sample

Because item difficulty and discrimination indices are dependent on the sample of students being tested, and in order to ensure that the MPT-G would be appropriate for students entering

a diverse array of courses, we asked participants in the working group to help solicit a number of pretest sites. With this assistance, we were able to test students at both two- and four-year schools, on the east and west sides of the state, and in general math[8] as well as precalculus courses. As shown in Tables 4 and 5, we tested a total of 2,642 students in 95 classes. Excluding forty-six students who did not mark the test version, arrived late to the testing session, or completed fewer than fifteen items, the final sample of 2,596 included 1,352 (52.1%) general math and 1,244 (47.9%) precalculus students.

Table 4. Number of CLASSES tested at each pretest site

Institution	General Math	Precalculus	Total
Eastern Washington University	17	6	23
Pierce College	3	3	6
Seattle Central Community College	1	0	1
Spokane Community College	2	2	4
Spokane Falls Community College	4	5	9
University of Washington-Bothell	0	2	2
University of Washington-Seattle	16	18	34
Whatcom Community College	2	6	8
Yakima Valley Community College	8	0	8
TOTAL	53	42	95

Table 5. Number of STUDENTS included in final dataset

Institution	General Math	Precalculus	Total
Eastern Washington University	485	191	676
Pierce College	88	83	171
Seattle Central Community College	8	0	8
Spokane Community College	32	53	85
Spokane Falls Community College	107	122	229
University of Washington-Bothell	0	42	42
University of Washington-Seattle	502	630	1,132
Whatcom Community College	9	140	149
Yakima Valley Community College	147	0	147
Total Tested	1,378	1,261	2,639
Excluded	26	17	43
TOTAL	1,352	1,244	2,596

Administration

Testing was conducted between September 25 and October 8, 2007. Test dates were staggered across institutions to allow OEA staff members to be onsite to coordinate testing and, when possible, administer the tests. Due to the number of simultaneous administrations, tests were often administered by the course instructor. Tests generally were given in the classroom in a single class period, although in one case they were administered in the college testing center over the course of several days. Instructions for administration are shown in Appendix B. All ten test versions were randomly assigned to students in each session. The number of usable tests of each version is shown in Table 6.

[8] Titles of courses included under General Math included Algebra Applied to Business and Economics; Algebra with Applications; Business Mathematics; Business Statistics; College Algebra; Finite Math, Functions, Models, and Quantitative Reasoning; Math for Elementary Education; Math for Liberal Arts; Nature of Math; and Survey of Math.

Table 6. Number of valid pre-tests by test version frequency (and percentage) and course type

Test	Version No.	General Math		Precalculus		Total	
Number Sense A	001	144	(10.7)	129	(10.4)	273	(10.5)
Number Sense B	002	139	(10.3)	126	(10.1)	265	(10.2)
Geometry A	003	140	(10.4)	118	(9.5)	258	(9.9)
Geometry B	004	145	(10.7)	120	(9.6)	265	(10.2)
Prob & Stats A	005	141	(10.4)	125	(10.0)	266	(10.2)
Prob & Stats B	006	138	(10.2)	122	(9.8)	260	(10.0)
Algebra A	007	127	(9.4)	126	(10.1)	253	(9.7)
Algebra B	008	130	(9.6)	127	(10.2)	257	(9.9)
Functions A	009	127	(9.4)	126	(10.1)	253	(9.7)
Functions B	010	121	(8.9)	125	(10.0)	246	(9.5)
TOTAL		1,352	(100.0)	1,244	(100.0)	2,596	(100.0)

Pretest Results

The purpose of the pretest was to identify high quality items to build parallel forms of the MPT-G and to begin populating a large item bank. Benchmark items were included to enable us to relate new MPT items to existing items and tests. We began our analyses of pretest data by examining the internal consistency of the benchmark total score and its ability to discriminate between students in different types of courses, and between those who did well on the full test and those who did less well. We followed this with analyses of individual benchmark items. Similar analyses were then carried out with respect to content (CRMT) test versions and items. For all analyses, omitted items were coded as incorrect. Table 7 shows the number of omissions on the Benchmark and CRMT items; in general, students omitted very few items.

Table 7. Number of items omitted for benchmark and CRMT tests

	Number of Omits					Number of Omits			
Test	n	Mean	Min.	Max.	Test	n	Mean	Min.	Max.
Benchmark	2596	0.2	0	4	Prob & Stats A	266	0.6	0	12
					Prob & Stats B	260	0.9	0	13
Number Sense A	273	0.8	0	12	Algebra A	253	1.9	0	13
Number Sense B	265	0.9	0	14	Algebra B	257	2.0	0	12
Geometry A	258	1.6	0	13	Functions A	253	1.6	0	13
Geometry B	265	1.8	0	15	Functions B	246	1.6	0	13

Benchmark Test and Items

Summary of findings. The internal consistency of the six-item Benchmark test was moderate, but was predicted to be high for an extended 30-item version. The total test score differentiated between general math and precalculus students, and was not influenced by the surrounding CRMT content (Number Sense, etc.). Each of the benchmark items discriminated both between general math and precalculus students and between low and high total test scores within each type of class. Thus, the six-item Benchmark test showed adequate reliability and validity.

Internal consistency. The internal consistency estimate of the Benchmark test using Cronbach's *alpha* was a moderate 0.64. The magnitude of this coefficient was slightly less than optimal, but was likely due to the small number of items. Based on the Spearman-Brown prediction formula, a test of 30 items, sampled from the same universe of items as those used, would have

achieved $\alpha \cong .90$. All items contributed to the consistency of test. As shown in Table 8, deleting any one of the items yielded a lower estimate of internal consistency than that observed for the test as a whole.

Table 8. Internal consistency of benchmark if items are removed

Item	Cronbach's α if Item Deleted
Item 13	.54
Item 14	.58
Item 15	.60
Item 16	.55
Item 17	.58
Item 18	.60

Note. n = 2596.

Difficulty and discrimination. To determine how well the Benchmark test differentiated between students who knew the relevant content and those who did not, we first carried out an analysis of variance to test for differences in benchmark total score by CRMT content and course type. The main effect of CRMT content was not statistically significant, $F(4,2586) = .02$, indicating that performance on the Benchmark test was independent of the surrounding CRMT content (Number Sense, etc.). The main effect of course type was large (*Mean*s = 4.4 vs. 3.5) and significant ($F(1,2586) = 227.8$, $p < .00001$, $d = .59$) with precalculus students outperforming general math students. The test of the interaction between CRMT content and course type was not significant.

Table 9. Descriptive statistics for benchmark items by course type

	Overall (n=2596)		General Math (n=1352)		PreCalc (n=1244)		Test for difference		
Score	Mn	SD	Mn	SD	Mn	SD	$F(1,2594)$	p	d
Item 13	.70	.46	.58	.49	.82	.38	184.0	1.5E-40	-0.54
Item 14	.49	.50	.42	.49	.56	.50	57.9	3.8E-14	-0.28
Item 15	.62	.48	.57	.50	.68	.46	36.3	2.0E-09	-0.23
Item 16	.65	.48	.57	.50	.74	.44	79.4	9.5E-19	-0.36
Item 17	.85	.36	.79	.41	.91	.28	78.6	1.4E-18	-0.34
Item 18	.61	.49	.54	.50	.68	.46	56.5	7.7E-14	-0.29
Total	3.91	1.63	3.47	1.70	4.40	1.41			

We next examined whether scores on the benchmark items discriminated between high and low performers *within* the two course types by computing correlations (r_{IT}) between each item score and the sum of the remaining items. We found all benchmark items to be at least fair discriminators within both groups (Table 10). The items tended to be slightly better at discriminating within general math classes than within precalculus classes, however, this difference was significant ($\alpha = .05$) only for items 13 and 15.

Table 10. Discrimination indices (corrected item-total correlations) for benchmark items by course type

	Overall (*n*=2596)		General Math (*n*=1352)		PreCalc (*n*=1244)		Test for difference	
Item	r_{IT}	*Interp.*	r_{IT}	*Interp.*	r_{IT}	*Interp.*	*z*	*p*
Item 13	.44	Good	.42	Good	.34	Fair	2.46	.01
Item 14	.39	Good	.34	Good	.29	Fair	1.28	.20
Item 15	.36	Good	.36	Good	.27	Fair	2.56	.01
Item 16	.41	Good	.40	Good	.34	Good	1.60	.11
Item 17	.35	Good	.32	Good	.27	Fair	1.37	.17
Item 18	.29	Fair	.27	Fair	.22	Fair	1.56	.12

CRMT Content Tests

Summary of findings. Comparison of student response to versions A and B of each CRMT content test (e.g., Number Sense, Algebra) showed no evidence of order effect; consequently these data were combined to form a single version of each content test for all subsequent analyses. The internal consistency of the content tests ranged from acceptable to very good, and the tests were of varying levels of difficulty. In general, students obtained highest scores on the Probability & Statistics test and lowest scores on Geometry and Functions. Precalculus students scored higher than general math student on all tests. Total scores on the CRMT tests were moderately related to total scores on the Benchmark test.

Order effects. As described above, we created two test versions for each of the five Content Standards to enable us to check and, if necessary, correct for possible order effects. Table 11 shows a variety of descriptive statistics for each of the resulting ten tests. To determine whether there were significant order effects, we compared mean CRMT total scores by test version for each of the five content areas. None of the *F*-tests were significant at $\alpha = .01$, nor did the mean number of items omitted differ by version. We therefore combined versions A and B within each content area for all subsequent analyses.

Table 11. Descriptives for each of the five CRMT tests by version

Test	*n*	*Mn*	SD	Min	Max	Mn Omits
Number Sense A	273	13.9	4.15	4	22	.8
Number Sense B	265	13.6	4.10	1	22	.9
Number Sense	538	13.8	4.13	1	22	.8
Geometry A	258	11.0	4.98	1	24	1.6
Geometry B	265	10.9	5.08	1	24	1.8
Geometry	523	10.9	5.02	1	24	1.7
Prob & Stats A	266	15.6	3.75	4	23	.6
Prob & Stats B	260	15.2	4.25	0	24	.9
Prob & Stats	526	15.4	4.01	0	24	.8
Algebra A	253	12.1	4.97	1	24	1.9
Algebra B	257	11.6	4.49	2	23	2.0
Algebra	510	11.8	4.74	1	24	2.0
Functions A	253	11.6	4.37	2	22	1.6
Functions B	246	11.7	3.85	4	21	1.6
Functions	499	11.6	4.12	2	22	1.6

Test reliability. We computed internal consistency (Cronbach's *alpha)* estimates for each of the five combined CRMT tests. As shown in Table 12, the reliability estimate was acceptable to very good for each test (range = .72 - .82).[9]

Table 12. Reliability coefficients for five content tests

CRMT Content	alpha	$alpha_a$	$alpha_b$	*n* cases
Number Sense	.76	.57	.63	409
Geometry	.82	.68	.71	351
Prob & Stats	.72	.53	.56	434
Algebra	.82	.66	.69	315
Functions	.74	.59	.56	321

Note. $alpha_a$ *and* $alpha_b$ refer to internal consistency estimates for first 12 and last 12 items, respectively.

Difficulty and discrimination. An analysis of variance showed that the five combined CRMT tests differed in overall difficulty, $F(4,2586) = 96.2$, $p = 2.2E\text{-}76$. A series of Tukey's HSD post-hoc tests identified the significant differences in means. As shown in Figure 1, the Geometry and Functions tests were the most difficult, and their means were not statistically significant from one another ('A'). Similarly, the means for Functions and Algebra were not significantly different from one another ('B') but were significantly lower than the mean for Number Sense ('C') which in turn was significantly lower than that of the easiest test, Probability & Statistics ('D').

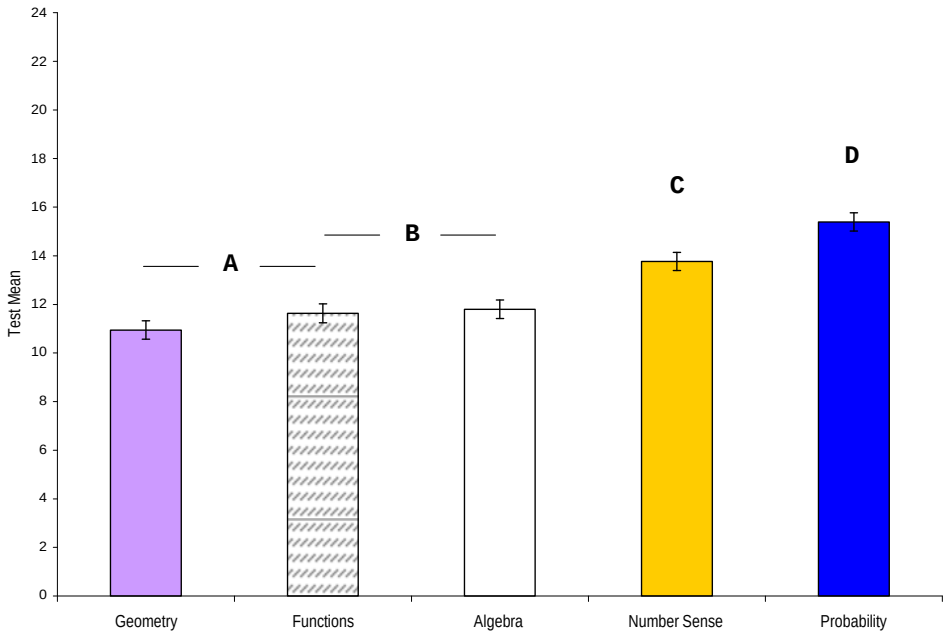


Figure 1. Mean test scores by content area

We also tested for mean differences between general math and precalculus students. As shown in Figure 2, precalculus student outperformed general math students on all tests (all *F*-values > 11).

[9] The number of cases reported in Table12 is lower than that shown in Table 11 due to the exclusion of tests with missing data.

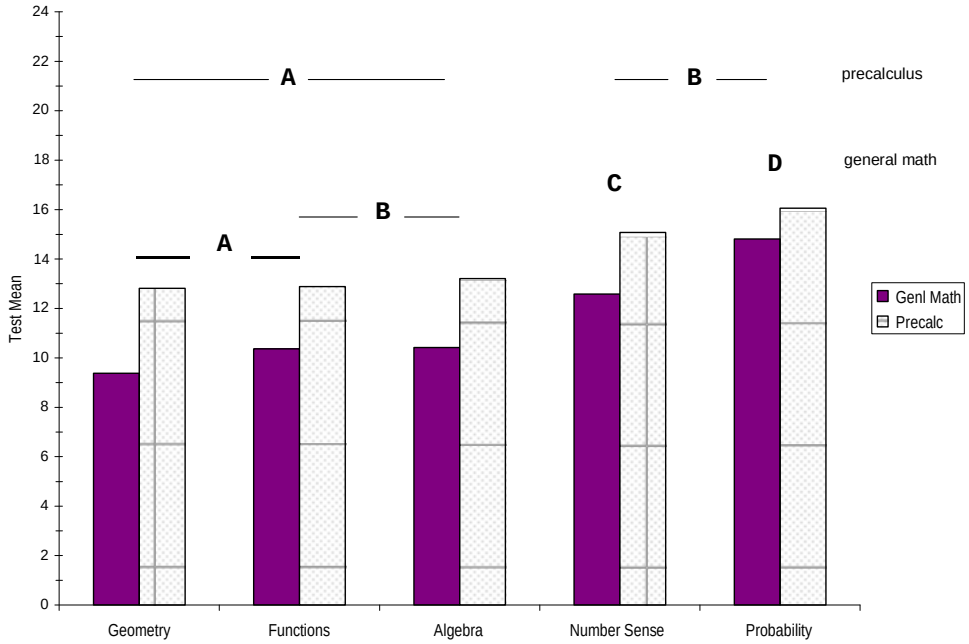


Figure 2. Mean test scores by content area and course type

Although the order of difficulty by content area was the same for general math and precalculus students, the pattern of significant differences was not. Specifically, for general math there were four homogeneous subsets (from most to least difficult): a) Geometry and Functions, b) Functions and Algebra, c) Number Sense, and d) Probability / Statistics. For precalculus, there were only two subsets: a) Geometry, Functions, and Algebra, and b) Number Sense and Probability & Statistics.

Relationship between CRMT tests and benchmark scores. Table 13 shows the correlations between the CRMT content test scores and the benchmark total score. Correlations ranged from .41 to .64, with the lowest value obtained for Probability and Statistics. This is as expected, given that the item pool from which benchmark items were drawn did not include this content area. After correction for attenuation, the correlations between the CRMT and Benchmark tests increased to .60 to .95 as shown in the table.

Table 13. Correlations between CRMT and benchmark total scores

CRMT Content	*r*	*r'*	*n*
Number Sense	.58	.85	538
Geometry	.56	.78	523
Prob & Stats	.41	.60	526
Algebra	.64	.95	510
Functions	.60	.88	499

Note. *r'* is the correlation between CRMT and benchmark score after correction for attenuation.

CRMT Content Items

Summary of findings. Each test contained some items that were not strongly related to overall test performance, did not predict performance on the benchmark test, or were too difficult or too easy. Overall, Geometry and Algebra were the strongest sets of items, and Probability &

Statistics was the weakest set. Most items contained two to three good alternatives (i.e., the correct answer and one or two good distractors).

Item difficulty and discrimination. We examined item discrimination through a series of analyses. We first conducted mixed between-within repeated measures analyses of variance on item means. The results revealed statistically significant mean differences between general math and precalculus students on zero (Probability & Statistics) to eighteen (Geometry) items (Table 14). Probability & Statistics items were less likely to discriminate because all students performed well on those items. Nonetheless, it appears at this time that the two groups of students were nearly equal in their knowledge of this content.

Using the sets of items for which there were significant mean differences by class type, we conducted stepwise discriminant function analyses to predict class membership. Table 14 shows that classification of students was very good for the four content areas other than Probability & Statistics.

Table 14. CRMT items that did and did not discriminate between general math and precalculus students

	Group means significantly different		DFA	
Content	Yes	No	X^2	% correctly classified
Number Sense	2, 3, 4, 5, 10, 12, 21, 22, 24, 25, 30	1, 6, 7, 8, 9, 11, 19, 20, 23, 26, 27, 28, 30	51.2	65.0%
Geometry	1, 3, 4, 6, 7, 9, 10, 12, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30	2, 5, 8, 11, 20, 23	79.7	69.0%
Probability/Stats		1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30	N/A	N/A
Algebra	4, 7, 10, 12, 21, 23, 25, 27, 30	1, 2, 3, 5, 6, 8, 9, 11, 19, 20, 22, 24, 26, 28, 29	35.0	64.5%
Functions	2, 9, 10, 19, 20, 23, 27, 28	1, 3, 4, 5, 6, 7, 8, 11, 12, 23, 24, 25, 26, 29, 30	63.2	69.7%

Note. To avoid capitalizing on chance, the alpha-level for each repeated measures ANOVA was set to $\alpha = .002$.

A second way in which we examined item discrimination efficacy was by computing item-rest correlations (r_{IT}) as a basis for classifying items as good, fair, or poor discriminators. Only three items were poor discriminators: Number Sense 6 and Functions 11 and 29. As shown in Table 15, the number of items within each content area that were classified as good discriminators ranged from nine (Probability & Statistics) to twenty (Geometry). Most (fifteen) of the Probability& Statistics items were classified as fair discriminators. In the cases of Number Sense and Functions, the numbers of items classified as fair and good were almost equal. Most (seventeen) of the Algebra items were classified as good discriminators.

Table 15. Number of CRMT items meeting discrimination and difficulty criteria

	r_{IT}	Hard 0 - 50%	Medium 50 - 85%	Easy 85 – 100%
		Number Sense		
Poor	< 0.1	1	0	0
Fair	0.1 - 0.3	6	5	0
Good	> 0.3	1	10	1
		Geometry		
Poor	< 0.1	0	0	0
Fair	0.1 - 0.3	3	1	0
Good	> 0.3	16	4	0
		Probability and Statistics		
Poor	< 0.1	0	0	0
Fair	0.1 - 0.3	4	10	1
Good	> 0.3	2	4	3
		Algebra		
Poor	< 0.1	0	0	0
Fair	0.1 - 0.3	4	3	0
Good	> 0.3	7	10	0
		Functions		
Poor	< 0.1	2	0	0
Fair	0.1 - 0.3	5	4	2
Good	> 0.3	6	4	1

We tested for differences in item-rest correlation coefficients by course type (general math vs. precalculus). There were only four significant differences: Number Sense 19, Probability & Statistics 21 and 30, and Algebra 10. Except for Algebra 10, in each case the item was stronger a discriminator among general math students than among precalculus students.

Relationships between CRMT item scores and benchmark total score. We tested items' abilities to predict benchmark total score using stepwise linear regression. A shown in Table 16, Number Sense, Geometry, and Algebra items produced the largest number of predictor items, and Probability& Statistics and Functions produced the smallest number. Nevertheless, the seven Functions items accounted for 40% of the variance in benchmark total score. The best predictor set was Algebra items, which accounted for 50% of the benchmark variance.

Table 16. Results of stepwise linear regression analyses predicting benchmark total score from CRMT item scores

	Num. Sense		Geometry		Prob/Stats		Algebra		Functions	
Item	B	*t*(526)	B	*t*(512)	B	*t*(518)	B	*t*(497)	B	*t*(491)
Item 1	.42	3.50								
Item 2					.52	3.00				
Item 3									.82	5.64
Item 4	.33	2.77					.44	3.85	.36	2.88
Item 5					.50	2.84				
Item 6							.40	3.58		
Item 7	.07	4.53			.28	2.03	.51	4.43	.50	3.96
Item 8							.23	2.13		
Item 9			.50	3.61					.46	3.77
Item 10	.73	4.92								
Item 11							.25	2.32		
Item 12	.32	2.49	.36	2.87			.67	6.14		
Item 19			.57	4.33					.77	6.43
Item 20	.27	2.32	.48	2.88			.27	2.48	.64	5.42
Item 21	.48	3.43	.32	2.50	.74	2.56	.23	2.16	.35	2.58
Item 22	.35	2.57								
Item 23					.33	2.03	.21	2.05		
Item 24	.30	2.50	.28	2.16			.80	7.15		
Item 25			.45	3.30	.56	3.70	.34	3.31		
Item 26			.44	3.28						
Item 27			.29	2.10						
Item 28	.38	2.34			.45	3.22	.38	2.72		
Item 29			.27	2.14						
Item 30	.60	5.11								
Constant	.96	4.99	1.88	10.66	1.64	5.65	1.48	11.11	1.85	14.03
R^2		.40		.33		.18		.50		.40

Distractor analysis. The process for the analysis of answer options was as follows. First, we examined the selection rate. An option had to have been selected by at least 5% of the respondents; otherwise, it was marked for discard. Second, for each content test we sorted respondents into quintiles based on their total test scores. Then, for each option, we conducted a one-sample chi-square test to determine whether the proportion of students choosing that alternative varied by quintile group. Alternatives for which the chi-square test was not significant were marked for discard because they did not discriminate among groups. Third, for each item, we plotted trace lines for each of the five alternatives (see Appendix E) and then visually examined each line to determine whether it was either monotonically decreasing or monotonically increasing. Trace lines for correct answers that were monotonically increasing (indicating that the probability of selecting that option was directly related to overall test performance) were marked as acceptable. Distractors (incorrect answers) with trace lines that were not monotonically decreasing were marked for revision or discard. Figure 3 shows frequency distributions of the number of items that had five, four, three, two, or one functional response alternatives. Figure 3 also includes the mean number of functional alternatives per item within each content area. In general, items contained two to three good alternatives (i.e., the correct answer and one or two good distractors).

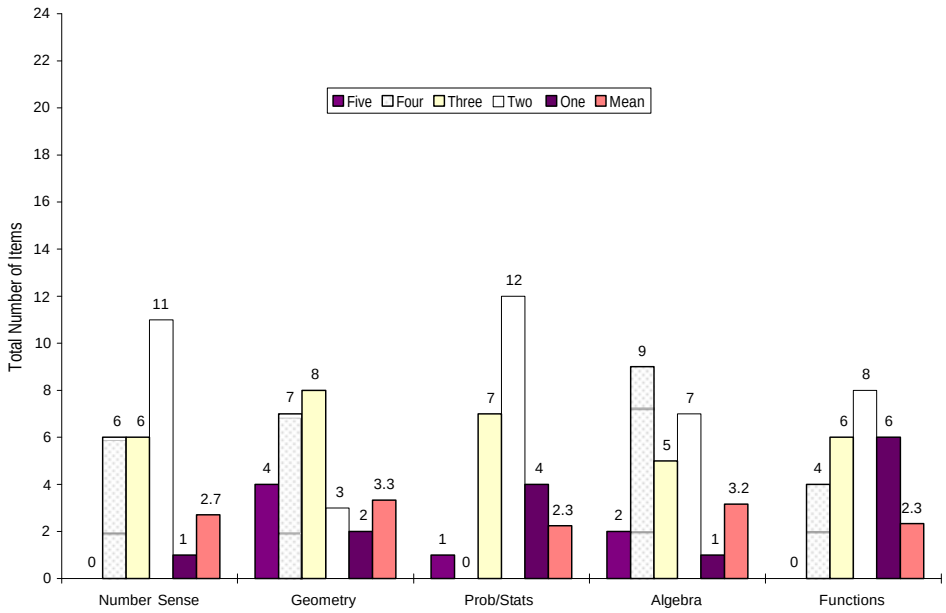


Figure 3. Number of items having varying numbers of functional response options by content area

INSTRUMENT DEVELOPMENT AND TESTING

The item-level pretest was carried out to determine item quality and to suggest ways in which individual items could be improved. Based on the pretest results, we selected 70 items to create two parallel forms of the MPT-G and carried out a pilot test to determine their reliability and predictive validity.

Pilot Test Methodology

Instruments

We compiled two parallel sets of 35 items and designated them test Versions I and J. Whereas each pretest instrument was made up of items from a single college readiness Standard, pilot test instruments were assembled by selecting items from across all five Standards. Each test consisted of seven Number Sense, six Geometry, six Probability & Statistics, eight Algebra, and eight Functions items, corresponding approximately to the proportions suggested in Table 1. Items were ordered throughout the test by alternating content areas, and balancing item difficulty from the beginning of the test to the end (Appendix F).

Because pretest analyses established that the majority of the CRMT items did not have four or even three functioning distractors, we decided to reduce the number of response alternatives utilized on the MPT-G. To establish the optimum item format, we created a 3-option form and a 4-option form of each item by dropping the poorest performing distractors (i.e., those that did not discriminate between high- and low-performing students or that were not selected by at least 5% of the Autumn 2007 test-takers).

As we had done on the pretest, we counterbalanced the order of the items to control for order effects. The resulting two (test version) by two (item format) by two (item order) design yielded a total of eight test versions.

Sample

Fewer institutions were able to participate in the pilot test than had been the case for the pretest and this, combined with lower math enrollments in winter quarter, resulted in a smaller sample size. Nevertheless we were able to administer the tests to students at a variety of institution types and in both general math and precalculus courses. As shown in Tables 17 and 18, we tested a total of 1,793 students in 69 classes. Excluding 227 students who did not mark the test version, arrived late to the testing session, or completed fewer than eighteen items, the final sample of 1,566 included 934 (59.6%) general math and 632 (40.4%) precalculus students.

Table 17. Number of CLASSES tested at each pilot test site

Institution	General Math	Precalculus	Total
Eastern Washington University	15	5	20
Spokane Falls Community College	4	4	8
University of Washington-Seattle	12	8	20
Whatcom Community College	3	6	9
Yakima Valley Community College	4	0	4
TOTAL	38	23	61

Table 18. Number of STUDENTS included in final dataset

Institution	General Math	Precalculus	Total
Eastern Washington University	415	159	574
Spokane Falls Community College	99	91	190
University of Washington-Seattle	401	298	699
Whatcom Community College	20	99	119
Yakima Valley Community College	44	0	44
Total Tested	979	647	1626
Excluded	45	15	60
TOTAL	934	632	1,566

Administration

Pilot tests were administered between January 8 and January 16, 2008. Test dates were staggered to allow OEA staff members to assist with test administration as they had during the pretest. Instructions for administration are shown in Appendix G. All eight test versions were randomly assigned to students in each session and the number of usable tests of each version is shown in Table 19.

Table 19. Number of valid pilot tests by test version frequency (and percentage) and course type

Test	Version No.	General Math		Precalculus		Total	
Version I, order 1, 3-option	001	116	12.4	87	13.8	203	13.0
4- option	002	121	13.0	78	12.3	199	12.7
order 2, 3- option	003	114	12.2	78	12.3	192	12.3
4- option	004	114	12.2	78	12.3	192	12.3
Version J, order 1, 3- option	005	120	12.8	73	11.6	193	12.3
4- option	006	112	12.0	85	13.4	197	12.6
order 2, 3- option	007	107	11.5	76	12.0	183	11.7
4- option	008	130	13.9	77	12.2	207	13.2
TOTAL		934	(100.0)	632	(100.0)	1,566	(100.0)

Pilot Test Results

Summary of findings. Each of the test version-form combinations demonstrated adequate internal consistency. Overall, students performed slightly better on test Version J than on test Version I, higher scores were obtained on 3-option tests than on 4-option tests, and precalculus students attained higher total scores than did general math students. The vast majority of the 70 items were found to be good or fair discriminators. Test scores were also very good predictors of course grades.

Test characteristics. There were usable tests from 934 general math students and 632 pre-calculus students. Table 20 shows internal consistency estimates (as measured by Cronbach's alpha) and mean scores by student group, test form, and test version. Each of the test version-form combinations demonstrated adequate internal consistency, and neither internal consistency nor mean score was affected by item order.

Table 20. Internal consistency estimates and mean scores by test version, item format, and course type

	Version I		Version J	
	3-option	4-option	3-option	4-option
Cronbach's α	.77	.75	.74	.75
General Math				
n	230	235	227	242
Mean (SD)	16.8 (5.8)	15.5 (5.3)	18.0 (5.2)	15.8 (5.6)
Mean %	47.9	44.4	51.5	45.0
Precalculus				
n	165	156	149	162
Mean (SD)	19.5 (5.1)	17.6 (5.4)	19.8 (5.3)	18.5 (5.4)
Mean %	55.6	50.3	56.6	52.8
Total				
n	395	391	376	404
Mean (SD)	17.9 (5.7)	16.4 (5.4)	18.7 (5.3)	16.9 (5.7)
Mean %	51.2	46.7	53.5	48.2

Note. The commonly accepted minimum standard for Cronbach's alpha is $\alpha = .70$.

Overall, students performed slightly better on test Version J than on test Version I (17.8 vs. 17.1, $F(1,1558) = 5.91$, $p = .02$, $d = .12$). The size of this effect was small in absolute terms and smaller than the effects of item format or course type. Pairwise comparisons revealed significant differences (at $p < .0014$) by test version on ten pairs of items. For six items, test-takers scored

higher on the J version (items 5, 13, 17, 19, 20, and 30), and for four items test-takers scores higher on the I version (items 10, 16, 18, 28).

On average, higher scores were obtained on 3-option tests than on 4-option tests (18.3 vs. 16.6, $F(1,1558) = 36.3$, $p = 2.1\text{-}9$, $d = -.30$), and precalculus students attained higher total scores than did general math students (18.8 vs. 16.5, $F(1,1558) = 69.4$, $p = 1.7\text{-}16$, $d = .43$). These effects are graphed in Figure 4. Note that students had only an estimated 40 minutes to complete the exam, rather than the customary 60 minutes, so scores likely are underestimates of students' "true" scores.

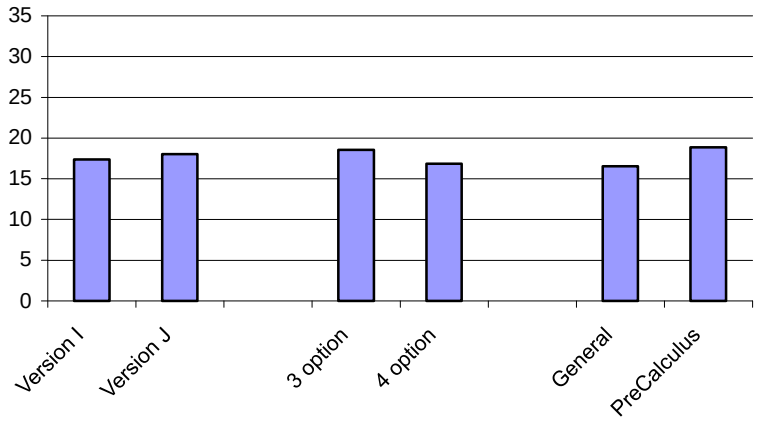


Figure 4. Mean scores by test version, item format, and course type

As shown in Figure 5, mean performance varied across the five CRMS content areas. Specifically, on average, students scored highest on Number Sense and lowest on Geometry and Functions. The profile of means followed very closely the mean item difficulties computed from the Autumn 2007 pretest data. However, separate analysis of the content means for the 3-option and 4-option groups revealed that the pattern only held for the 3-option group (Figure 6). Surprisingly, the 4-option test appears to have produced a floor effect: All five means were very similar to one another and only one mean was greater than 50%.

One possible explanation for the poorer performance on the 4-option test versions is that, at 35 items, the test was too long: To test this, we computed separate scores for each of five sets of seven items (i.e., items 1 through 7, 8-14, etc.). Repeated measures analyses of these segment-scores showed that group performance differences were evident as early as the second segment (Figure 7). Although scores in both groups tended to decline over the course of the test, the decline was much steeper among those in the 4-option group. Thus, the floor effect was not caused by poor performance only in the latter stages of the test.

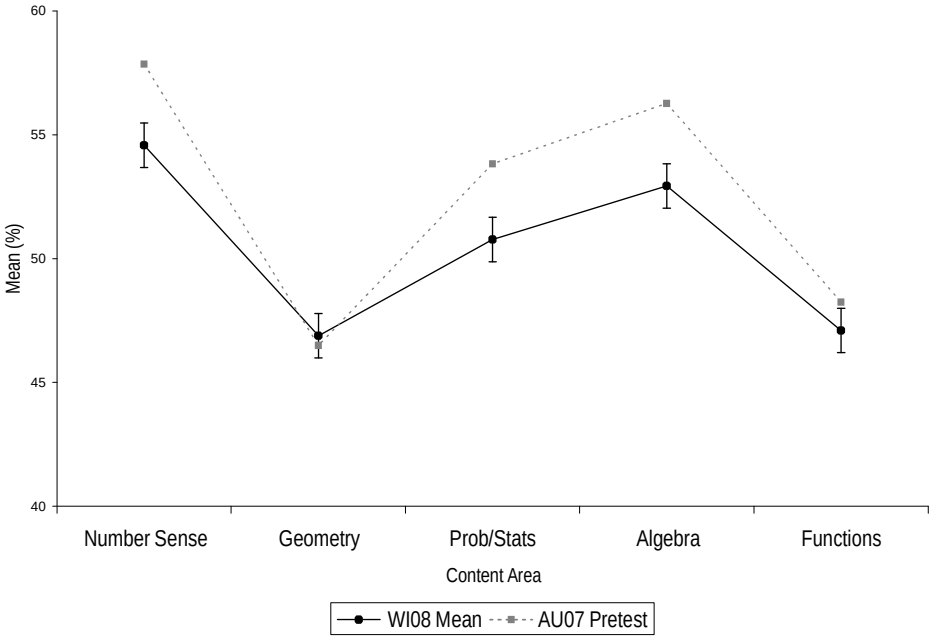


Figure 5. Mean scores (% correct) by CRMS content area

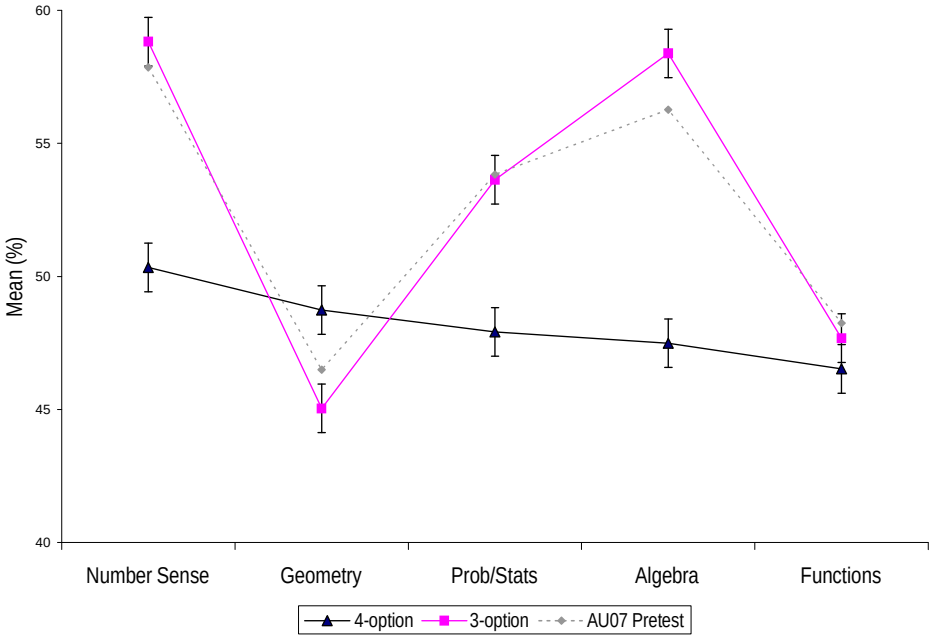


Figure 6. Mean scores (% correct) by CRMS content area and number of response options

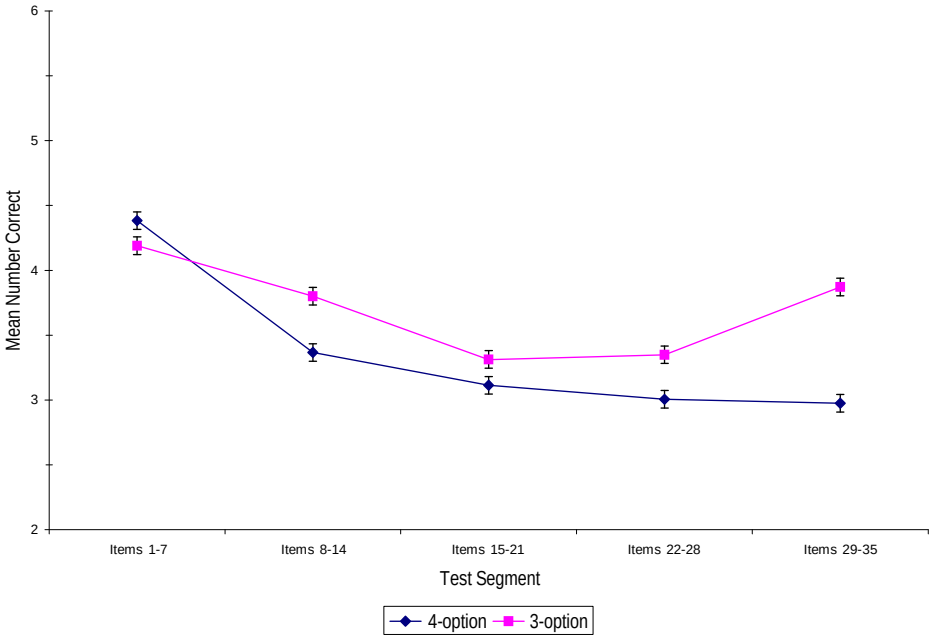


Figure 7. Mean scores (% correct) by test segment and number of response options

Discrimination. We computed a discrimination index for each item as the point-biserial correlation between scores on that item and scores on the total test. The vast majority of items were good ($r > .3$) or fair ($r = .1$-$.3$) discriminators. Only two Version I items were poor discriminators (16, 18), as were two Version J items (18, 23).

Distractor analysis. Distractor analyses were carried out in a manner similar to that described for the pretest, and revealed that items tended to have about two functioning distractors. A functioning distractor was defined as one which a) was selected by at least 5% of all test-takers and b) discriminated between high- and low-performing test-takers. As shown in Table 21, only a minority of the 4-option items had three functioning distractors. The apparent superiority of a 3-option item format has been supported by other researchers in the field of test development.[10]

Table 21. Results of distractor analyses

	Version I		Version J	
	3-option	4-option	3-option	4-option
Average number of good distractors	1.57	1.66	1.43	2.06
Number of items with 3 good distractors	--	4	--	9
Number of items with 2 good distractors	21	16	17	17
Number of items with 1 good distractors	12	13	15	8
Number of items with 0 good distractors	2	2	3	1

[10] See particularly: Trevisan, M.S., Sax, G., & Michael, W.B. (1991). The effects of the number of options per item and student ability on test validity and reliability. *Educational and Psychological Measurement*, *51*, 829-837; Rodriguez, M.C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement, Issues and Practice*, *24*, 3-13.

Predictive and Convergent Validity. To examine the predictive validity of test scores, we looked at the relationship between scores and test-takers' final grades in the math class in which they took the pilot test. Final grades were available for 1,437 of the students. Of those, 33 received a W grade (withdrew), leaving 1,404 cases for further analysis. Grades from Yakima Valley Community College and Whatcom Community College were converted from letters to numbers (A = 3.95, B = 3.0, C = 2.0, etc.). The average final grade was 2.60 (SD = 1.1), but the average grade in general math (M = 2.67) exceeded the average grade in precalculus (M = 2.48), $F(1,1402) = 9.86$, $p = .002$.

The zero-order correlation between course grade and test score was $r = .32$. The value of the coefficient did not change when we controlled for version (I vs. J) and number of answer options. However, the correlation was stronger within general math classes ($r = .38$) than within precalculus classes ($r = .29$). In other words, the MPT-G score was a slightly better predictor of performance in non-precalculus courses.

We used logistic regression to predict course success or failure. We defined success in two ways: 1) a final grade equal to or greater than 2.0 (C) or 2) a grade equal to or greater than 2.5 (C+/B-). Approximately three-fourths of all students achieved a final grade of at least 2.0 (Table 22).

Table 22. CRMT mean scores by course grade and class type

	General			Precalculus		
Course Grade	n	Proportion of cases	Mean MPT-G score (SD)	n	Proportion of cases	Mean MPT-G score (SD)
2.0 or higher	692	.80	17.4 (5.5)	400	.74	19.7 (5.3)
2.5 or higher	564	.65	17.8 (5.6)	308	.57	20.3 (5.4)
3.0 or higher	412	.48	18.7 (5.6)	221	.41	20.7 (5.5)

Analyses were conducted using the entire sample and then separately for the general math and precalculus groups. Thus, we tested six models of the form:

$$\log_e\left(\frac{p_i}{1-p_i}\right) = Intercept + Slope * MPTGscore_i$$

where p_i was the probability of success in the math course.

The data in Table 23 show that MPT-G total test score was a statistically significant predictor of course success for both general math and precalculus. Note, however, that due to the negative skew of course grades, the logistic model was much better for classifying "successful" cases than "unsuccessful" cases. This was particularly true under the more lenient definition of success (i.e., grade ≥ 2.0).

Table 23. Results of logistic regressions predicting course success from CRMT test score

							% Correctly Classified	
Criterion and Class Type	Omnibus χ^2	Intercept	*Slope*	Wald χ^2	*p*	Odds Ratio	Fail	Success
Course grade ≥ 2.0								
General Math	62.05	-.68	.13	53.51	3E-13	1.14	2.9	99.6
Precalculus	24.68	-.66	.09	23.41	1E-06	1.10	3.6	99.0
All cases	70.80	-.45	.10	65.08	7E-16	1.11	1.6	99.9
Course grade ≥ 2.5								
General Math	84.09	-1.45	.13	72.41	2E-17	1.14	27.1	86.2
Precalculus	39.61	-1.66	.10	33.97	6E-09	1.11	40.2	74.7
All cases	99.30	-1.28	.10	90.43	2E-21	1.11	28.0	84.9

The odds ratios derived from the logistic regression analyses indicated that each 1-point increase in MPT-G total score was associated with a 1.1 – 1.4 (or 2-3%) increase in the odds of course success. Using those ratios, we calculated – for each class type -- estimates of the MPT-G total scores associated with particular probabilities of success. As shown in Figure 8, based on the pilot test sample data, about 76% of students with an MPT-G score of 14 (out of 35) would achieve a grade of 2.0 in general math, but only about 66% of such students would achieve a grade of 2.0 in precalculus. In order to have a 75% probability of grade of 2.0 in precalculus, a student would need an MPT-G total score of 19.

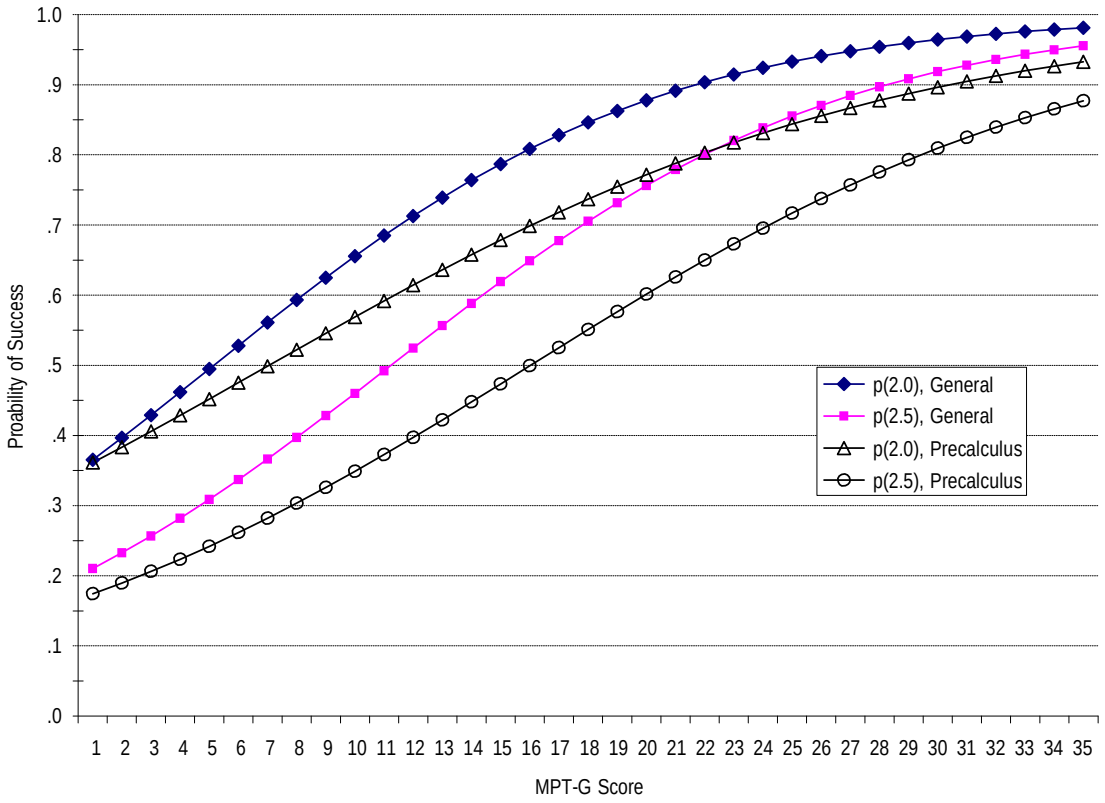


Figure 8. Expected probabilities of success in math courses for MPT-G scores.

A word of caution about the results of the logistic regression is in order. Recall that students in the pilot test did not have the full60 minutes to complete the test and, because of this, it is very likely that their scores were depressed. Thus, the derived model may underestimate probability of success in general math.

APPENDICES

Appendix A. Item Layouts for Pretest Versions 001-010

Number Sense

Test Item No.	Test Version 001	Test Version 002
1	4.1 real numbers	Version 001 item no. 25
2	4.1 real numbers	Version 001 item no. 26
3	4.2 computation	Version 001 item no. 27
4	4.2 computation	Version 001 item no. 28
5	4.2 computation	Version 001 item no. 29
6	4.3 estimation strategies	Version 001 item no. 30
7	4.1 real numbers	Version 001 item no. 19
8	4.2 computation	Version 001 item no. 20
9	4.2 computation	Version 001 item no. 21
10	4.2 computation	Version 001 item no. 22
11	4.3 estimation strategies	Version 001 item no. 23
12	4.3 estimation strategies	Version 001 item no. 24
13	MPT-I benchmark 1	Version 001 item no. 13
14	MPT-I benchmark 2	Version 001 item no. 14
15	MPT-I benchmark 3	Version 001 item no. 15
16	MPT-I benchmark 4	Version 001 item no. 16
17	MPT-I benchmark 5	Version 001 item no. 17
18	MPT-I benchmark 6	Version 001 item no. 18
19	(parallel to item 1)	Version 001 item no. 7
20	(parallel to item 2)	Version 001 item no. 8
21	(parallel to item 3	Version 001 item no. 9
22	(parallel to item 4)	Version 001 item no. 10
23	(parallel to item 5)	Version 001 item no. 11
24	(parallel to item 6)	Version 001 item no. 12
25	(parallel to item 7)	Version 001 item no. 1
26	(parallel to item 8)	Version 001 item no. 2
27	(parallel to item 9)	Version 001 item no. 3
28	(parallel to item 10)	Version 001 item no. 4
29	(parallel to item 11)	Version 001 item no. 5
30	(parallel to item 12)	Version 001 item no. 6

Geometry

Test Item No.	Test Version 003	Test Version 004
1	5.2 represent situations	Version 003 item no. 25
2	5.2 represent situations	Version 003 item no. 26
3	5.3 draw and justify conclusions	Version 003 item no. 27
4	5.3 draw and justify conclusions	Version 003 item no. 28
5	5.3 draw and justify conclusions	Version 003 item no. 29
6	5.4 apply trigonometric relationships	Version 003 item no. 30
7	5.2 represent situations	Version 003 item no. 19
8	5.3 draw and justify conclusions	Version 003 item no. 20
9	5.3 draw and justify conclusions	Version 003 item no. 21
10	5.3 draw and justify conclusions	Version 003 item no. 22
11	5.3 draw and justify conclusions	Version 003 item no. 23
12	5.4 apply trigonometric relationships	Version 003 item no. 24
13-30	… follows Test Versions 001 and 002	

Probability and Statistics

Test Item No.	Test Version 005	Test Version 006
1	6.1 solve problems	Version 005 item no. 25
2	6.1 solve problems	Version 005 item no. 26
3	6.2 develop tables and plots	Version 005 item no. 27
4	6.2 develop tables and plots	Version 005 item no. 28
5	6.3 develop inferences and predictions	Version 005 item no. 29
6	6.4 create and evaluate	Version 005 item no. 30
7	6.1 solve problems	Version 005 item no. 19
8	6.2 develop tables and plots	Version 005 item no. 20
9	6.2 develop tables and plots	Version 005 item no. 21
10	6.3 develop inferences and predictions	Version 005 item no. 22
11	6.3 develop inferences and predictions	Version 005 item no. 23
12	6.4 create and evaluate	Version 005 item no. 24
13-30	… follows Test Versions 001 and 002	

Algebra

Test Item No.	Test Version 007	Test Version 008
1	7.1 recognize and use	Version 007 item no. 25
2	7.1 recognize and use	Version 007 item no. 26
3	7.2 combine and simplify	Version 007 item no. 27
4	7.3 solve	Version 007 item no. 28
5	7.3 solve	Version 007 item no. 29
6	7.3 solve	Version 007 item no. 30
7	7.1 recognize and use	Version 007 item no. 19
8	7.1 recognize and use	Version 007 item no. 20
9	7.2 combine and simplify	Version 007 item no. 21
10	7.3 solve	Version 007 item no. 22
11	7.3 solve	Version 007 item no. 23
12	7.3 solve	Version 007 item no. 24
13-30	… follows Test Versions 001 and 002	

Functions

Test Item No.	Test Version 009	Test Version 010
1	8.1 recognize relationships	Version 009 item no. 25
2	8.1 recognize relationships	Version 009 item no. 26
3	8.2 represent functions	Version 009 item no. 27
4	8.2 represent functions	Version 009 item no. 28
5	8.3 analyze/interpret features	Version 009 item no. 29
6	8.4 model relationships	Version 009 item no. 30
7	8.1 recognize relationships	Version 009 item no. 19
8	8.1 recognize relationships	Version 009 item no. 20
9	8.2 represent functions	Version 009 item no. 21
10	8.3 analyze/interpret features	Version 009 item no. 22
11	8.4 model relationships	Version 009 item no. 23
12	8.4 model relationships	Version 009 item no. 24
13-30	… follows Test Versions 001 and 002	

Appendix B. Instructions to Administer Math Pretest

Please read the following instructions through before beginning the test administration.

1) **Count all test booklets**. Enter the number and your initials on the Test Count Verification form.

2) Ask students to be seated and quiet. Read the following aloud:

Before we begin today's test, I have been asked to read the following directions. The purpose of this test is to create a new math placement test to be used in this course. For this reason, results on the tests will be compared to final course grades for the class as a whole. Your individual results will be confidential, and this test will not affect your course grade in any way. Participation is voluntary. However, the test will give both you and me a better idea of your level of mathematical preparation for this course, so it is important that you do your best on it.

In a moment, I will pass out the test materials. Please clear your desk. For this test, you may NOT use calculators, calculator watches, laptops, or other similar aids. You must use a #2 pencil to mark on the answer sheet: I have extras if you need one. I will also be providing scratch paper which you must return at the end of the period. You are NOT allowed to use your own scratch paper. I will now pass out the testing materials. Do not open the test booklet or make any marks on the answer sheet until you are told to do so.

3) Hand out one test booklet, answer sheet, and two pieces of scratch paper to each student. Read the following aloud:

On Side 1 of the answer sheet, print the name of your math course and your section designation above the "Name" box, then print and bubble in your full name.

Next, print and bubble in the test version in columns A-C under "Identification Number." The test version is printed on the top right hand corner of your test booklet. Be sure to enter the leading zeros. For instance, for test version 009 you would bubble in 0, 0, 9.

Finally, print and bubble in your student number in columns J-P.

You do not need to fill in any of the other identification boxes.

Please do NOT make any marks in the test booklet. If you find that your test booklet has been written on, raise your hand and we will give you another copy.

You will have 50 minutes to complete the test. Be sure that you answer all 30 items carefully. If you finish early, check over your work and remain quietly seated.

Are there any questions?

You may open your test booklets and begin work now.

4) Five minutes before the end of the class period, read the following aloud:

Please stop work immediately, put your pencil down, and close your test booklet. Double-check that you have filled in your name, student number, and test version. Remain seated while I collect a test booklet from each of you. After I have collected all booklets, hand me your answer sheet, scratch paper, and #2 pencil, if I provided you one, as you leave the classroom.

5) Collect materials as described above. **Count all test booklets** and enter the count and your initials on the Test Count Verification form. Put the completed answer sheets and used scratch paper in the envelope marked with the appropriate section.

*** Because this is a secure test, it is very important that all test booklets be accounted for. ***

Appendix C. Pretest Item Discrimination by Test Content

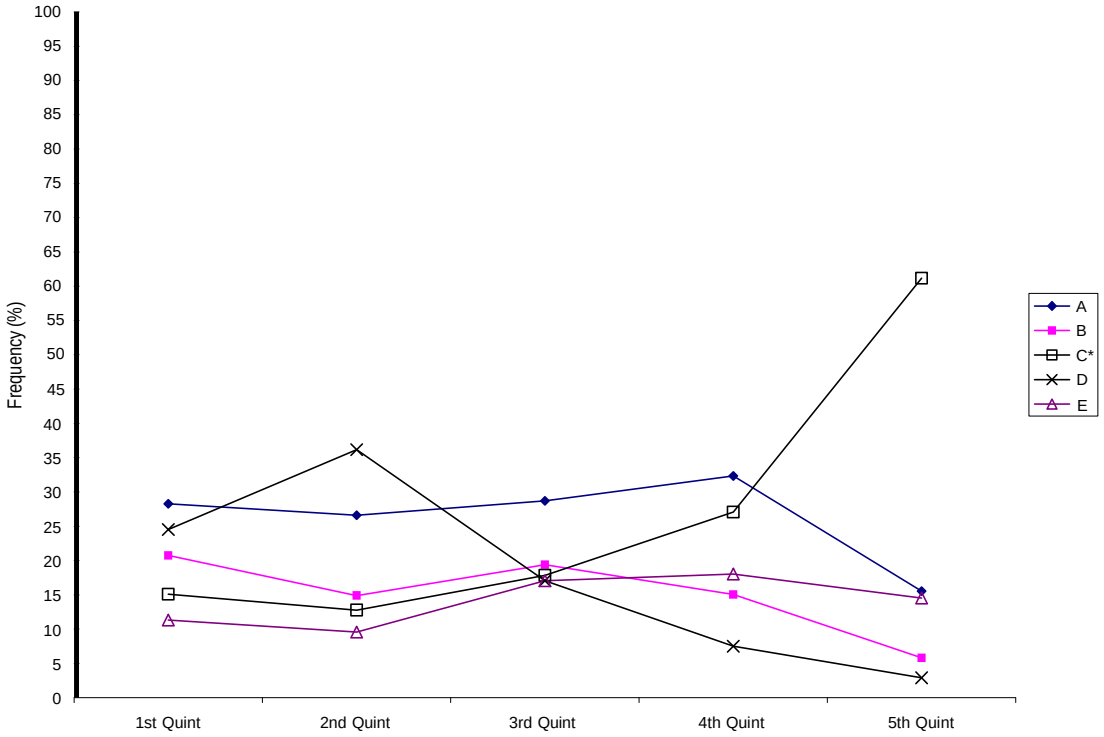
	Num Sense		Geometry		Prob Stat		Algebra		Functions	
Item	r_{IT}	D	r_{IT}	D	r_{IT}	D	r_{IT}	D	r_{IT}	D
1	.17	.29	.49	.69	.23	.40	.32	.35	.20	.26
2	.28	.46	.30	.36	.40	.43	.35	.53	.36	.62
3	.42	.61	.38	.52	.38	.32	.34	.56	.43	.63
4	.38	.53	.25	.29	.29	.43	.40	.61	.44	.70
5	.25	.33	.33	.44	.33	.39	.27	.44	.21	.30
6	-.01	.10	.36	.52	.24	.40	.37	.57	.19	.34
7	.25	.40	.33	.41	.23	.37	.27	.46	.32	.47
8	.33	.36	.32	.45	.23	.26	.21	.34	.25	.52
9	.30	.36	.39	.59	.27	.39	.36	.55	.42	.64
10	.34	.40	.48	.72	.28	.45	.23	.34	.31	.33
11	.19	.32	.31	.51	.21	.33	.37	.62	.03	.05
12	.39	.53	.40	.57	.28	.40	.40	.60	.20	.30
19	.22	.30	.49	.71	.27	.47	.31	.29	.31	.55
20	.20	.36	.20	.26	.32	.40	.41	.60	.30	.54
21	.39	.44	.46	.61	.36	.20	.36	.54	.40	.61
22	.22	.28	.20	.28	.34	.56	.32	.43	.37	.63
23	.18	.27	.44	.60	.35	.47	.30	.48	.22	.39
24	.35	.52	.38	.56	.27	.38	.39	.58	.28	.41
25	.29	.37	.22	.25	.33	.49	.19	.35	.22	.34
26	.36	.41	.39	.58	.31	.26	.27	.48	.29	.49
27	.30	.27	.35	.53	.22	.19	.39	.56	.42	.63
28	.34	.33	.42	.61	.28	.44	.36	.37	.29	.29
29	.29	.41	.37	.55	.18	.34	.28	.42	.05	.09
30	.46	.64	.38	.52	.22	.36	.43	.65	.25	.31
N	538		523		526		510		499	

Note. r_{IT} is the correlation between an item score and the sum of the remaining items. D is the Index of Discrimination and was computed by subtracting the mean among those who total score was in the lowest 27th percentile from the mean among those who total score was in the highest 27th percentile.

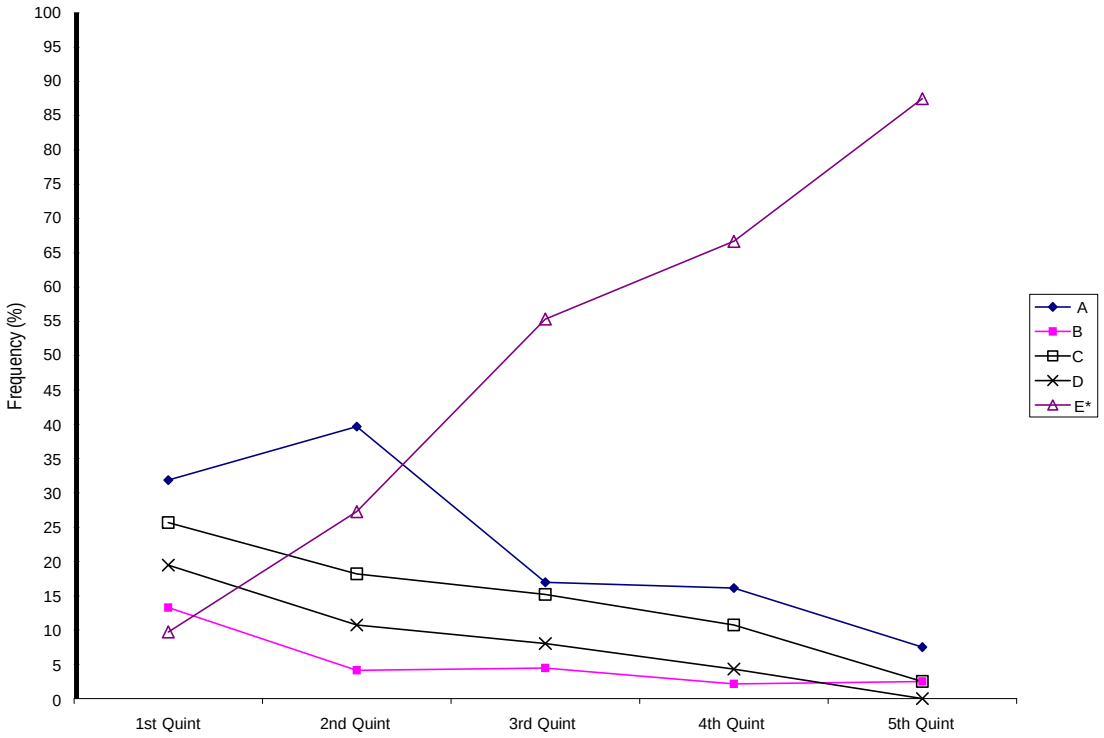
Appendix D. Pretest Item Difficulty by Test Content

	Num Sense		Geometry		Prob Stat		Algebra		Functions	
Item	Mn	SD	Mn	SD	Mn	SD	Mn	SD	Mn	SD
1	.66	.47	.41	.49	.44	.50	.18	.38	.85	.36
2	.61	.49	.80	.40	.79	.41	.47	.50	.43	.50
3	.61	.49	.40	.49	.87	.34	.53	.50	.70	.46
4	.44	.50	.21	.41	.31	.46	.45	.50	.44	.50
5	.26	.44	.28	.45	.79	.41	.49	.50	.20	.40
6	.21	.41	.45	.50	.50	.50	.59	.49	.71	.46
7	.60	.49	.25	.43	.53	.50	.71	.45	.32	.47
8	.78	.41	.72	.45	.84	.37	.35	.48	.50	.50
9	.80	.40	.52	.50	.71	.45	.65	.48	.47	.50
10	.78	.41	.46	.50	.52	.50	.28	.45	.86	.35
11	.30	.46	.48	.50	.62	.49	.52	.50	.08	.28
12	.64	.48	.46	.50	.72	.45	.58	.49	.22	.42
19	.73	.44	.48	.50	.40	.49	.17	.37	.56	.50
20	.47	.50	.84	.36	.75	.44	.55	.50	.52	.50
21	.75	.43	.45	.50	.94	.24	.43	.50	.65	.48
22	.23	.42	.34	.47	.47	.50	.78	.41	.49	.50
23	.27	.44	.41	.49	.74	.44	.48	.50	.43	.50
24	.60	.49	.38	.49	.69	.46	.58	.49	.77	.42
25	.74	.44	.26	.44	.32	.47	.54	.50	.27	.45
26	.79	.41	.66	.47	.89	.32	.47	.50	.58	.49
27	.86	.35	.41	.49	.89	.31	.60	.49	.41	.49
28	.83	.37	.46	.50	.42	.49	.18	.38	.86	.35
29	.28	.45	.43	.50	.55	.50	.68	.47	.13	.34
30	.52	.50	.36	.48	.70	.46	.54	.50	.18	.38
N	538		523		526		510		499	

Appendix E. Sample Pretest Distractor Analysis Graphs



Number Sense 5



Functions 21

Appendix F. Pilot Test Item Difficulty by Test Form

		Form I (N=786)		Form J (N=780)	
Item	Content	Mn	SD	Mn	SD
1	Num Sense	.60	.49	.63	.48
2	Geometry	.64	.48	.58	.49
3	Prob/Stat	.74	.44	.71	.45
4	Algebra	.52	.50	.54	.50
5	Functions	.62	.49	.78	.42
6	Num Sense	.59	.49	.54	.50
7	Geometry	.49	.50	.51	.50
8	Prob/Stat	.56	.50	.59	.49
9	Algebra	.60	.49	.53	.50
10	Functions	.72	.45	.60	.49
11	Num Sense	.41	.49	.43	.50
12	Algebra	.54	.50	.59	.49
13	Prob/Stat	.23	.42	.43	.50
14	Algebra	.42	.49	.42	.49
15	Functions	.52	.50	.58	.49
16	Num Sense	.34	.47	.26	.44
17	Geometry	.44	.50	.54	.50
18	Prob/Stat	.49	.50	.36	.48
19	Algebra	.49	.50	.72	.45
20	Functions	.20	.40	.32	.47
21	Num Sense	.56	.50	.51	.50
22	Geometry	.37	.48	.42	.49
23	Functions	.27	.45	.22	.41
24	Algebra	.52	.50	.58	.49
25	Functions	.39	.49	.43	.49
26	Num Sense	.69	.46	.72	.45
27	Geometry	.37	.48	.39	.49
28	Prob/Stat	.49	.50	.38	.48
29	Algebra	.49	.50	.47	.50
30	Functions	.36	.48	.44	.50
31	Num Sense	.59	.49	.64	.48
32	Geometry	.36	.48	.42	.49
33	Prob/Stat	.52	.50	.57	.50
34	Algebra	.49	.50	.42	.49
35	Functions	.49	.50	.48	.50
	Num Sense	.54	.24	.53	.23
	Geometry	.45	.26	.48	.25
	Prob/Stat	.50	.23	.51	.21
	Algebra	.51	.24	.53	.24
	Functions	.45	.20	.48	.21
	TOTAL	.49	.16	.51	.16

Appendix G. Instructions to Administer Math Pilot Test

1) Ask students to be seated and quiet. Read the following aloud:

Before we begin today's test, I have been asked to read the following directions. The purpose of this test is to create a new math placement test to be used in college level math courses. Your answer to each item will be compared to your final math grade to show how well the test works. However, your individual results will be confidential, and this test will not affect your course grade in any way. Participation is voluntary. However, the test will give us a better idea of your level of mathematical preparation, so it is important that you do your best on it.

Please clear your desk of all materials. For this test, you may NOT use calculators, calculator watches, laptops, or other similar aids. You must use a #2 pencil to mark on the answer sheet – I have extras if you need one. I will also be providing scratch paper which you will be asked to return at the end of the period. You are NOT allowed to use your own scratch paper. I will now pass out the testing materials. Please do not open the test booklet or make any marks on the answer sheet until you are told to do so.

2) Hand out one test booklet, answer sheet, and two pieces of scratch paper to each student. Read the following aloud:

On Side 1 of the answer sheet, print the name of your math course and your section designation above the "Name" box, then print and bubble in your full name.

Next, print and bubble in your student number under "Identification Number" starting in column "A".

Finally, print and bubble in the test version in columns N through P under "Special Code." The test version is printed on the top right hand corner of your test booklet. Be sure to enter the leading zeros. For instance, for test version 008 you would bubble in 0, 0, 8.

You do not need to fill in any of the other identification boxes. Please do NOT make any marks in the test booklet. If you find that your test booklet has been written on, raise your hand and we will give you another copy.

You will have until the end of the class period to complete the test. Be sure that you answer all 35 items carefully. The test is longer than usual, so don't worry if you don't finish. However, if you do finish early, check over your work and remain quietly seated. Are there any questions?

You may open your test booklets and begin work now.

3) Several minutes before the end of the class period, read the following aloud:

Please stop work immediately, put your pencil down, and close your test booklet. Turn over your answer sheet and double-check that you have filled in all the required information. Remain seated while I collect a test booklet, answer sheet, and scratch paper from each of you.

4) Collect materials as described above.

5) Put the completed answer sheets and used scratch paper in the envelope marked with the appropriate section. ►►Students who arrive late may be allowed to take the test. However, if a student begins the test more that 10 minutes late, please mark the "16" bubble in the "Grade or Educ" box on side 1 of the answer sheet.

***** Because this is a secure test, it is very important that all test booklets be accounted for. *****