Mary Garner, Meghan Burke, and Jose Binongo

Analysis of a College Placement Test in Mathematics Using the Rasch Measurement Model

Purpose/Objectives

The purpose of this paper is to describe how the Rasch Measurement Model was used to analyze the properties of a college placement test in mathematics. The placement test is used to make recommendations to college students regarding which of four different courses they should take: Precalculus, College Algebra, College Trigonometry, or Calculus. The Rasch model was used to confirm faculty assumptions about the difficulty of the items on the test and to determine whether student performance on the placement test can be used to predict success in the four mathematics courses.

Perspectives: The Placement Test

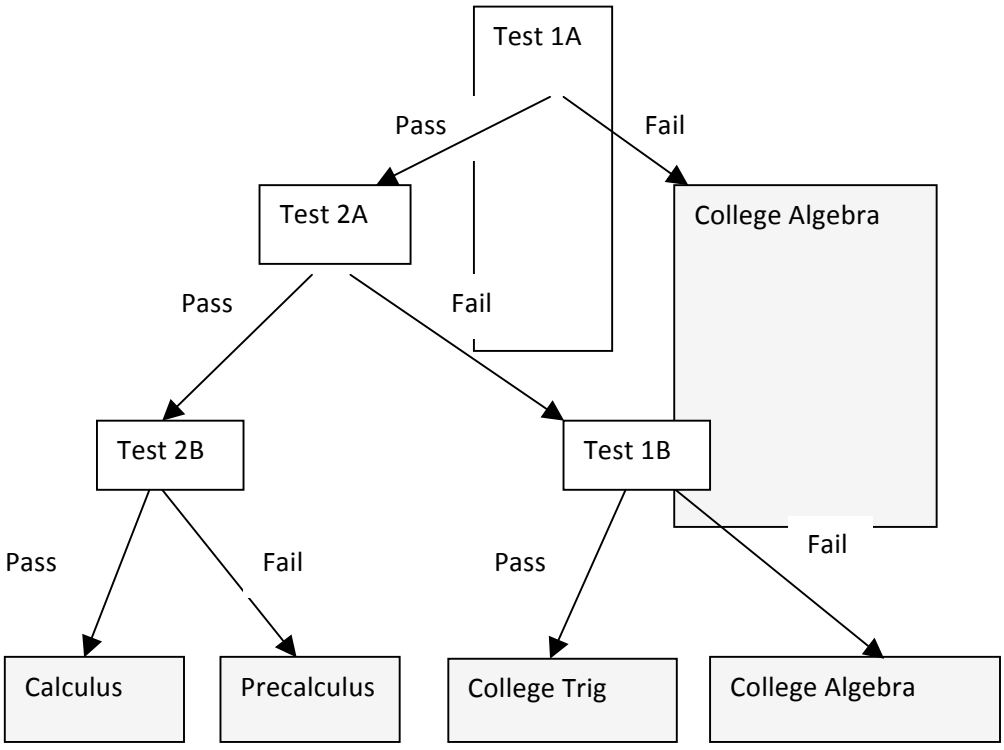
At Kennesaw State University, students who are majoring in mathematics or science are required to take a one-semester Precalculus course and a one-semester Calculus course as part of their core requirements. In Fall of 2003, out of 1,087 students enrolled in precalculus at KSU, 43% of the students received a D, F, W (withdraw), or WF (withdraw failing); and in Fall of 2004, that number was 36%. In Fall of 2003, out of 470 students enrolled in calculus 41% of the students received a D,F,W, or WF; and in the Fall of 2004, that number was 45%. In an effort to improve student performance in these courses, in Fall of 2005, the Department of Mathematics and Statistics started offering the students the option of taking Precalculus as a two-semester sequence instead of a one-semester course. The new two-semester sequence included College Algebra in the first semester and College Trigonometry in the second semester. In order to help students in their decision about whether to take the two-semester Precalculus sequence, the one-semester Precalculus, or go straight

into Calculus, a committee of faculty members was formed to design a placement test, named the Mathematics Placement Test (MAPT).

The MAPT consists of a total of 51 multiple choice items divided into four subtests.

- Test 1A: 15 simple algebra questions.
- Test 1B: 11 more complex algebra questions.
- Test 2A: 12 simple trigonometry questions
- Test 2B: 13 more complex trig questions.

A student passes any one of the tests if the student gets at least 60% of the items correct. Students navigate through the test as shown below by starting with Test 1A. If they fail Test 1A, they are placed in College Algebra. If they pass Test 1A, they take Test 2A. If they fail Test 2A, they take Test 1B. If they fail Test 1B, they are placed in College Algebra. If they pass 1A and 1B (but failed 2A), they are placed in College Trigonometry. If they pass 1A and 2A, they take 2B. If they pass 1A, 2A, and 2B, they are placed into Calculus. If they pass 1A and 2A but they fail 2B, they are placed in Precalculus.



None of the student takes the entire test of 51 items. The structure of the resulting data is shown below. All of the students take items 1 through 15. A portion of the students then take items 16-38 and another portion of the students take items 27-51.

Student Placements	**Test 1A Items 1-15**	**Test 1B Items 16-26**	**Test 2A Items 27-38**	**Test 2B Items 39-51**
College Algebra				
College Trigonometry				
Precalculus				
Calculus				

The placement test is not yet required of all students, nor are students required to follow the placement recommendation given at the end of the test. Consequently, there are students who might have taken an easier course than recommended or a more difficult course than recommended. In addition, students enrolled in a given course would have different instructors.

Methods

The Partial Credit Model (Masters, 1982) (PCM) is a mathematical model that belongs to the general class of models known as item response or latent trait models. Specifically, the PCM belongs to the family of Rasch measurement models (Wright & Stone, 1979) which can be viewed as the simplest of the item response models (Embretson & Hershberger, 1999). Using the PCM, a scale or ruler can be established along which we can simultaneously locate items according to difficulty and persons according to ability. The PCM describes the relationship between person ability and item difficulty probabilistically. The probability that person v will

receive a score of k on particular item i ($a_{vi} = k$), where the possible scores are 0 through m, is expressed in terms of the person's ability b_v, and the difficulty of going from a rating k-1 to a rating k on that item d_{ik} as follows:

$$\Pr(a_{vi} = k) = \frac{e^{\sum_{j=1}^{k} (b_v - d_{ij})}}{1 + \sum_{h=1}^{m} e^{\sum_{j=1}^{h} (b_v - d_{ij})}} \quad \text{for h= 1 to m,}$$

and

$$\Pr(a_{vi} = 0) = \frac{1}{1 + \sum_{h=1}^{m} e^{\sum_{j=1}^{h} (b_v - d_{ij})}} \quad \text{for h = 1 to m.}$$

Using FACETS software, a joint maximum likelihood method is employed for estimation of person and item parameters of the PCM. According to Linacre (1983), there is no need to impute vales for missing data as long as the "observations form a linked network such that every parameter can be estimated unambiguously within the same frame of reference." Since all students take items 1 through 15, the system would be adequately linked.

The person parameters represent person ability and the item parameters represent item difficulty, both of which can be placed on a common scale or ruler with measurements in "logits" or log odds units.

The PCM can accommodate both dichotomous and polytomous items. The items on the MAPT were scored dichotomously as right or wrong. The item difficulties for the MAPT were obtained in a first run of FACETS, then with those values anchored, FACETS was run again with the grade in the course treated as an item and scored polytomously as 1 for a F, 2 for a D, 3

for a C, 4 for a B, and 5 for an A. The probability of scoring a given grade could then be examined as a function of ability level.

Data Sources

A preliminary analysis was conducted on 1619 students from the 2005-2006 school year. The students who chose to take the placement test in Fall of 2005 and Spring of 2006 were used. Of the 1619 students, 355 took College Algebra, 13 took College Trigonometry, 275 took Precalculus and 43 took Calculus. Since there were so few cases for College Trigonometry and Calculus, performance in those courses was not examined. Another analysis is being planned with more extensive data and will be completed by April of 2010.

Preliminary Results

Table 1 shows for each item, the item difficulty, standard error, infit mean square statistic, outfit mean square statistic, and the subtest to which the item belonged. The items are tabulated in order of increasing difficulty. The same item difficulties are shown visually in Figures 1 and 2. The items from Test 1A clustered at the lower end of the scale, as expected. The items from Test 2B clustered at the higher end of the scale, also as expected. Interestingly, the items from Test 1B (more advanced algebra) and Test 2A (simple trigonometry) appeared to be equivalent in difficulty, clustering along the middle of the scale.

As shown in Figures 3 and 4, performance on the test did predict performance in College Algebra and Precalculus, with the probability of getting a higher grade in the course increasing with increasing ability as measured by the test. Clearly, the D and F categories should be collapsed since a D was never a most probable grade. A student was most likely to earn at least a C in College Algebra at an ability level of -.76; while a student was most likely to earn at least a

C in Precalculus at an ability level of -.56. As expected, it takes more ability to score at least a C in Precalculus than in College Algebra, but there is not as great a separation as expected. College Algebra is recommended for students who get 9 items correct on Test 1A; however, there are 13 items on Test 1A that fall below the level of -.76 so it may be that the first passing score is too low for allowing students to proceed to Test 2A. In addition, the students must then get 8 of the 12 items on Test 2A to be considered for Precalculus, but only 3 of the 12 items fall below the level of -.56; consequently, the passing score for Test 2A may be set too high.

Item	Test	Difficulty	SE	Infit	Outfit
4	1a	-2.19	.07	.93	.91
8	1a	-2.05	.07	1.08	1.88
2	1a	-2.03	.07	1.01	1.30
6	1a	-1.81	.07	1.07	1.22
9	1a	-1.68	.06	1.07	1.09
5	1a	-1.61	.06	.95	1.01
1	1a	-1.30	.06	1.05	1.10
29	2a	-1.21	.08	.95	1.12
7	1a	-.98	.06	1.00	1.05
3	1a	-.98	.06	1.04	1.06
36	2a	-.95	.07	.82	.72
10	1a	-.93	.06	1.05	1.17
14	1a	-.91	.06	1.00	1.08
12	1a	-.88	.06	.96	.97
11	1a	-.78	.06	1.01	1.00
24	1b	-.76	.1	.90	.89
31	2a	-.62	.07	.99	1.00
16	1b	-.59	.1	.94	.91
21	1b	-.46	.1	.96	.94
38	2a	-.33	.07	.87	.85
34	2a	-.26	.07	.92	.93
20	1b	-.22	.10	.96	.94
27	2a	-.2	.07	.94	.91
19	1b	-.19	.10	.95	.99
23	1b	-.11	.10	.95	.93
28	2a	-.10	.07	.97	.95
15	1a	.03	.06	1.11	1.14
33	2a	.09	.07	.94	.97
35	2a	.15	.07	.85	.84
32	2a	.25	.07	.93	.91
37	2a	.31	.07	.92	.89
18	1b	.33	.10	.92	.88
26	1b	.38	.10	.94	.89

22	1b	.5	.10	.94	.91
43	2b	.57	.10	.94	.90
25	1b	.67	.10	1.19	1.20
42	2b	.68	.10	.88	.84
17	1b	.69	.11	.96	.96
44	2b	.84	.10	.90	.85
30	2a	.89	.07	.95	.92
13	1a	1.05	.07	1.24	1.45
41	2b	1.15	.10	.97	.99
47	2b	1.20	.10	.96	.94
50	2b	1.42	.10	.88	.84
51	2b	1.47	.10	.93	.93
49	2b	1.74	.11	.92	.90
48	2b	1.8	.11	.95	.97
40	2b	1.87	.11	1.04	1.07
46	2b	1.96	.11	.95	.89
45	2b	2.34	.12	.90	.89
39	2b	2.62	.13	.99	1.00

Table 1. Item difficulty, standard error, infit mean square statistic, and outfit mean square statisitic for each item.

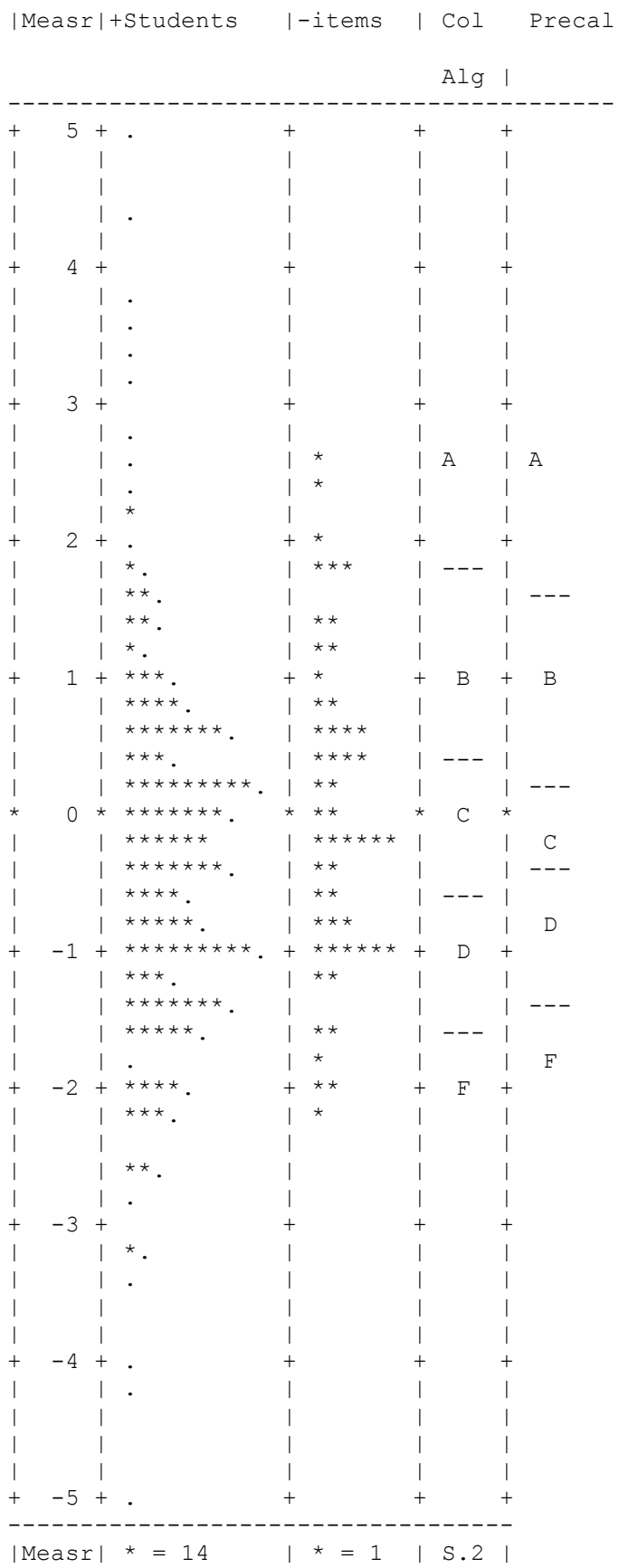


Figure 1. Wright map for items on the MAPT.

Figure 2. Distribution of item difficulties for each test.

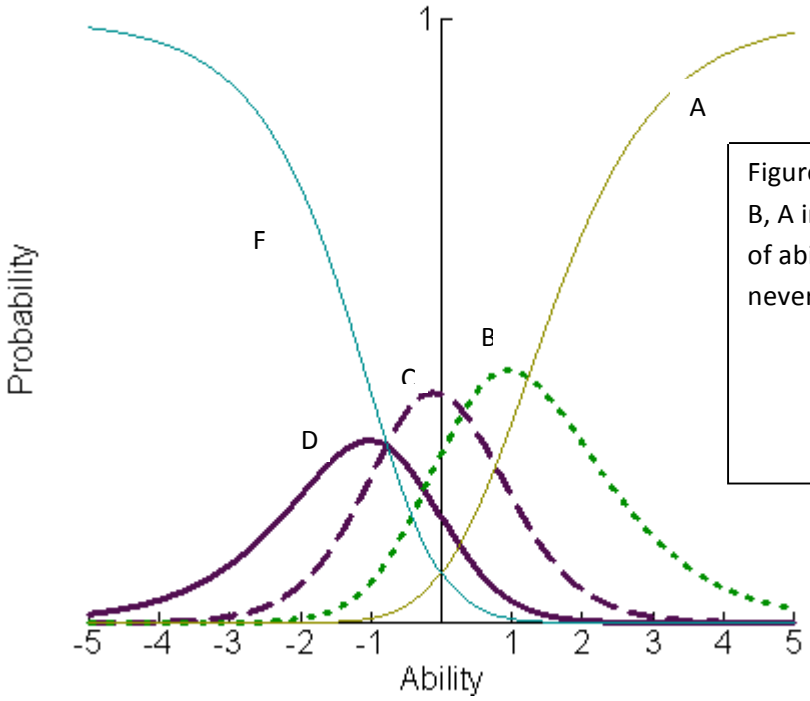


Figure 3. Probability of Scoring F, D, C, B, A in College Algebra as a function of ability in logits. (Note scoring a D is never most likely for any ability level.)

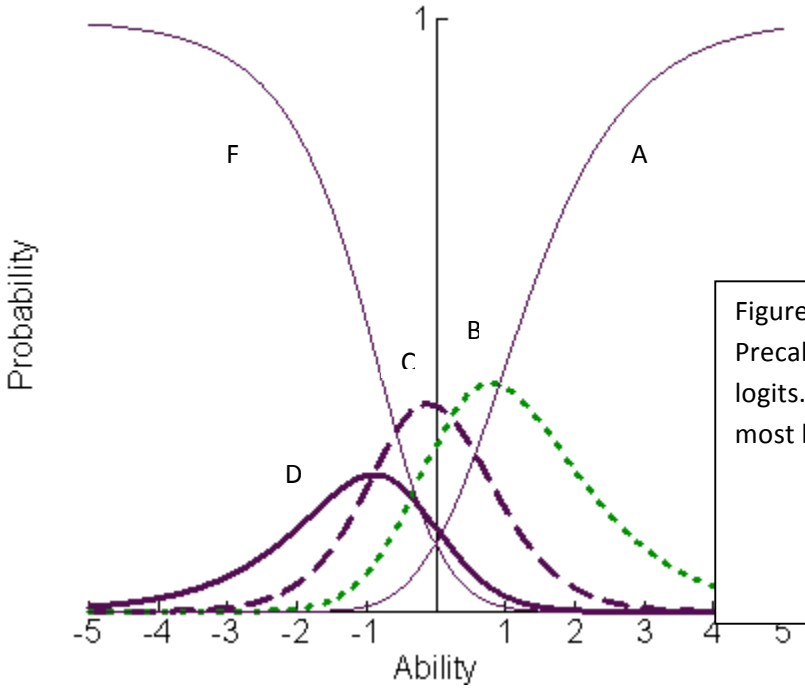


Figure 4. Probability of F, D, C, B, A in Precalculus as a function of ability in logits. (Note that scoring a D is never most likely for any ability level.)

Significance

` Much has been written about the usefulness of the Rasch Model for measuring cognitive development by allowing the placement of items and persons on a continuum that spans grade levels (Bond & Fox, 2007). Little has been written about applying the same technique to college course placement issues, but it seems natural to apply the model to a multi-level placement test such as the one described in this paper and establish a continuum that spans mathematical expertise and readiness for the next course. Applying the Rasch model to such a data set also provides an opportunity to examine the performance of the model in the presence of missing data, a much discussed issue in the Rasch literature.

Plans

By April 2010 we anticipate having a larger data set at our disposal and applying the same analysis as described above to the data for all four of the courses in which students may be placed. Enough data may be available to extend the analysis to include instructors of the courses as a facet in the model to compensate for differences in grading difficulty.

References

Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Social Sciences,* second edition.

Embretson, S.E. & Hershberger, S.L. (1999). *The New Rules of Measurement: What Every Psychologist and Educator Should Know.* Lawrence Erlbaum Associates.

Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47(2), 149-174.

Linacre, J. M. (1989). Many-facet Rasch measurement. Chicago: MESA Press.

Linacre, J. M., & Wright, B. D. (1992). Facets (Version 2.5) [Computer software]. Chicago: MESA Press.