

# Developing, Validating, and Using A Placement Test

**John W. Kenelly**  
Clemson University

**John G. Harvey**  
University of Wisconsin–Madison

**This paper discusses the creation of the placement tests found in the Maplesoft™-MAA Placement Test Suite. It reviews how the Mathematical Association of America (MAA) determined what subject areas required placement tests, how they developed new placement tests in the past, and how those tests were validated. The last section provides tips on using the Placement Test Suite, including information on setting cutoff scores for your institution.**

## Building and Validating a New Placement Test

The first step in developing a new placement test was determining what kind of new test was needed to ensure that colleges and universities could assign students to courses at an appropriate level. Discussions between the MAA Committee on Testing (COT) and collegiate faculty identified areas where the previous MAA placement tests did not suit college and university needs. For example, colleges and universities have multiple entry-level algebra courses and typically do not have an introductory geometry course. For that reason, the MAA Placement Test Program (PTP) included several algebra tests and none for geometry.

Once the need for a test in a specific area was identified, the next step was to create a test panel of college and university faculty. Members of the test panel were chosen based upon their experience in teaching the courses for which the placement test was to be developed. After a list of faculty prospects was identified by the COT, these faculty members were asked if they were interested in helping to develop the new placement test and if they would be able to serve on the test panel. The COT test panels were constituted of some members who were experienced test constructors and some who were new to test development. The experienced test constructors were drawn from among the members of previous COT panels and individuals who had extensive test development experience with the College Board and the Educational Testing Service.

Once a test panel was formed the first task was to formulate the overall specifications for entry into the designated course. This task required the panel members to determine what students need to know and should be able to do. This resulted in a set of content codes. The content code for an item is a description of the mathematical content to be tested. Examples of content codes are as follows:

- Linear equations
- Algebraic fractions
- Word problems; linear equations

*Each* question on *each* test in the Placement Test Suite has content codes that are accessible when an instructor examines a test in the Suite.

After the test panel completed its list of content codes, the members of the panel drafted items. The panel edited the items and used them to assemble a draft of the new test. Once the draft test was completed, the test panel, with the assistance of volunteer post-secondary institutions, gave the test to students in order to develop statistical information that provided both test data and item strips for individual questions. For a definition of item strips, see below.

The test data consisted of the following: (a) the mean score, (b) the standard deviation, and (d) the Cronbach  $\alpha$ . Since a placement test should discriminate over the full range of scores a mean of 50 percent is ideal. An unusual standard deviation will suggest that the distribution of difficulty levels is inappropriate. The Cronbach  $\alpha$  is a coefficient of reliability (consistency) determined by inter-item correlations. If the inter-item correlations are high, then the items are measuring the same underlying construct. This is what is meant by high or good reliability. Table 1 displays the test data for a sample test; Table 2 displays the distribution of scores for the same test. Figure 1 is a graph of the data displayed in Table 2.

Number of students	294
Number of test items	32
Mean score	20.31
Standard deviation	5.605
Reliability (coefficient $\alpha$ )	0.811

Table 1. Test summary statistics.

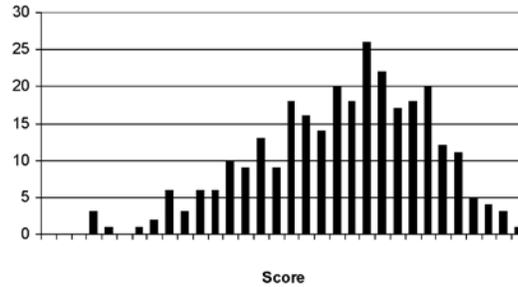


Figure 1. Graph of test score distribution.

Scores	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Frequency	0	0	0	3	1	0	1	2	6	3	6	6	10	9	13	9
Scores	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Frequency	18	16	14	20	18	26	22	17	18	20	12	11	5	4	3	1

Table 2. Test score distribution.

Table 3 and Figure 2 constitute an *item strip*. An item strip contains these data: (a) the number of students choosing each of the multiple-choice options within each quintile, (b) the difficulty level of the item (the percent of students who answer the item correctly) within each quintile, (c) the  $r$ -biserial statistic of the item (defined below), and (d) the quintile performance graph (Figure 2). Distracters should attract poorer students in equal proportions. The correct answer should be selected by the higher achieving students; otherwise, a flawed distracter (incorrect option) is improperly attracting able students. Thus, the quintile performance graph should be uniformly increasing to the right in the graph as in Figure 2; otherwise a weaker group of students is outperforming a group of better students.

The correct response to the item whose data is presented in Table 3 was choice E; thus, the probability that students will respond correctly to this item is 0.73, and the  $r$ -biserial for the item is 0.42. Here is the definition of the  $r$ -biserial statistic.

$r$ -biserial: A correlation coefficient relating performance on a test question and performance on the total test. It is an index of discrimination measuring the extent to which candidates who score high on the total test tend to get the question correct and those who score low tend to get the question incorrect.

	A	B	C	D	E	Invalid
Quintile						
First	3	12	13	4	23	3
Second	1	9	5	3	40	1
Third	0	7	3	4	45	0
Fourth	1	5	1	1	51	0
Fifth	0	2	0	1	56	0
Probability	0.02	0.12	0.07	0.04	[0.73]	0.01
$r$ -biserial	-0.15	-0.19	-0.07	-0.06	[0.42]	-0.11

Table 3. Matrix of responses by quintile.

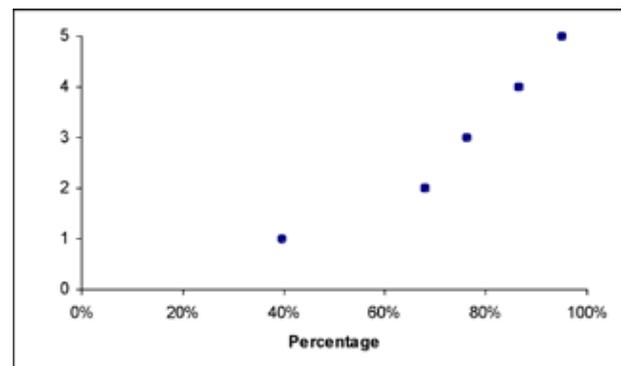


Figure 2. Percent responding correctly by quintile.

The item data and test statistics were used by the COT test panel to edit the questions and to assemble the next version of the test. This process was repeated until a satisfactory test was constructed and validated. At this point the test was released to the MAA Placement Test Program (PTP) for inclusion in the test packet. This addition to the test packet included two parallel forms of the developed test. As the test was used, additional parallel forms were produced to give subscribers several forms that they could use as equivalent tests. Since parallel forms is a debatable concept in the testing community, PTP subscribers were encouraged to collect data locally and calibrate their cutoff scores across the different forms.

The COT, through its *Placement Test Newsletter*, maintained close contact with the PTP subscribers. As time passed it would sometimes be clear that an existing test needed to be replaced with a significantly revised one. When this need arose a new test panel was created and the process of test development and validation was repeated. In particular, the Placement Test Suite includes Arithmetic and Skills Test, Forms 2A, 2B, 2D, and 2E. The “2” indicates that there has been a significant improvement of the Arithmetic and Skills Test (old version 1) and the inclusion of four new forms labeled A, B, D, and E.

If, as a user of the Placement Test Suite, you modify an existing test by deleting some items or by replacing them, pay close attention to the content codes and be prepared to collect data locally to validate your new test. If you shuffle the items or reorder the multiple choices within an item on an existing test, the same attention should be paid to validation as when you delete or replace an item.

## Using the Placement Test Suite

Once you have determined your placement test needs and purchased the Placement Test Suite, you must select the “best fit” test or tests in the Suite. You can allow students to progress through two or more tests to determine their best placement. To construct the best placement program your institution may want to work with an external consultant who is experienced in the development and installation of placement test programs. The MAA can provide a list of approved consultants with these credentials.

Cutoff scores are an essential feature of any placement testing program. These scores must be determined by studying and using locally generated data. There are many sources of information that can advise you about the development of appropriate cutoff scores, including some within the Maplesoft-MAA Placement Test Suite documentation. In general, the development of cutoff scores requires several major steps:

1. Collecting local data to be used to develop the initial cutoff scores
2. Setting the initial cutoff scores
3. Iteratively improving the cutoff scores using the data from your placement testing program

Most institutions like to use multiple forms of their placement tests in order to facilitate administration of the program. Multiple forms allow retesting, flexible scheduling, and improve test security. The Placement Test Suite provides four forms of each of the tests for you to use for this purpose. The four forms of each test were developed to the highest professional standards of test development. However, to ensure that your program best serves your students, we recommend that you calibrate scores across the different forms. Just as there are multiple references on setting cutoff scores, there are multiple references on calibrating parallel forms. We have found the following paper to be useful.

Buras, A. (January 1996) Test Equating Procedures: A Primer on the Logic and Applications of Test Equating. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA (ED 395 038).

Since the multiple forms in the Placement Test Suite were built to be parallel forms, the linear equating method is suggested. “The function relating scores on Form X to scores on Form Y using linear equating is generally expressed as  $Y^* = a(X - c) + d$ ; where ‘a’ refers to the ratio of the standard deviation of Form Y to the standard deviation of Form X, ‘c’ refers to the mean of Form X, and ‘d’ refers to the mean of Form Y.” (Buras, 1996, p. 6)

Placement testing programs are dynamic, that is, they require continual maintenance. Cutoff scores should be reexamined on a regular basis. The placement tests that are being used should be examined to determine if they are still appropriate.

## Conclusion

We hope that this document helps you make decisions regarding the placement tests from the Placement Test Suite that you choose. As was true for the MAA Placement Test Program the Mathematical Association of America is ready to help you as much as possible as you choose tests, engage in local validation of those tests, and calibrate equivalent forms of the tests so that your students are appropriately placed.



[www.maplesoft.com](http://www.maplesoft.com) | [info@maplesoft.com](mailto:info@maplesoft.com)