



**MAPP™**

*Listening. Learning. Leading.*

**Validity of the Measure of  
Academic Proficiency  
and Progress (MAPP™) Test**

**MEASURABLE RESULTS. IMMEASURABLE VALUE.**

A Summary from ETS

## **Validity of the Measure of Academic Proficiency and Progress (MAPP)**

John W. Young  
Center for Validity Research  
Educational Testing Service

### **Introduction**

This document describes the validity evidence for the *Measure of Academic Proficiency and Progress*<sup>™</sup> (MAPP<sup>™</sup>). MAPP is designed to assess general education student learning in two- and four-year colleges and universities in order to improve the quality of instruction and learning. MAPP is suitable for use as the outcome measure in a direct value-added assessment, which can be used to estimate institutional effects on student learning. MAPP was introduced in January 2006, and is the successor to Academic Profile. MAPP is an integrated test of four academic skills (critical thinking, reading, writing and mathematics), and measures these skills in three contexts: humanities, social sciences and natural sciences. MAPP comes in both paper-and-pencil and computer-delivered versions. MAPP comes in two forms: the Standard Form with 108 items and an Abbreviated Form in which each individual student responds to 36 of the 108 items. The Standard Form provides a total score, scaled subscores and proficiency classifications at both the group and individual student levels, while the Abbreviated Form reports reliable group data, but only a total score for individual students. MAPP has the same test format as Academic Profile, and scores from MAPP have been equated to the Academic Profile score scale so that MAPP scores can be compared to earlier results for Academic Profile.

### **Test Structure of MAPP**

MAPP measures four types of skills: One-fourth of the items measure critical thinking skills; one-fourth measure reading skills; one-fourth measure writing skills; and one-fourth measure mathematics skills. Each item measuring critical thinking skills or reading skills is associated

with a particular academic context: humanities, social sciences or natural sciences. Most of the reading and critical thinking items are based on information presented in a brief reading selection, a picture or a graph. Typically, a reading selection will serve as the stimulus for one or more reading items and one or more critical thinking items.

The following table illustrates the structure of the Standard Form:

Academic Context	Type of Skill			
	Critical Thinking	Reading	Writing	Mathematics
Humanities	9 items	9 items	27 items	27 items
Social Sciences	9 items	9 items		
Natural Sciences	9 items	9 items		
<b>Total Items</b>	<b>27 items</b>	<b>27 items</b>	<b>27 items</b>	<b>27 items</b>

MAPP focuses on the skills acquired through general education courses rather than on the knowledge acquired about the subjects taught. MAPP yields two types of information: norm-referenced scaled scores and criterion-referenced proficiency classifications. Eight scaled scores are computed and reported for each examinee taking the Standard Form:

1. a total score (108 items)
2. four skills subscores (27 items each): Critical Thinking, Reading, Writing and Mathematics
3. three context-based subscores (17 to 19 items each): Humanities, Social Sciences and Natural Sciences

In addition to the eight scaled scores, nine proficiency classifications are computed and reported for each examinee taking the Standard Form. The skill areas of writing and mathematics are each divided into three proficiency levels on the basis of the specific skills measured by the items. The skill area of reading is divided into two proficiency levels and combined with critical

thinking, which serves as the third proficiency level of this combined skill area. Each item on the MAPP test is associated with a particular proficiency level in a particular skill area, depending on the specific skill that it measures. Each student taking the MAPP test is classified as “Proficient,” “Marginal” or “Not proficient” at each proficiency level of each skill area. These nine proficiency classifications are reported for each examinee.

### **Validity**

For all testing programs, validity is the foremost quality concern. Validity is a judgment of the extent to which empirical evidence and scientific theories support the interpretations, inferences and actions based upon scores from a test (Messick, 1989). The validity of a test is neither a fixed entity nor an immutable quality, as validity depends on a number of factors, including the intended uses of the scores, the characteristics of the examinees and evidence accumulated over time. A test may be highly valid for one purpose, such as making admissions decisions, but have poor validity for another purpose, such as course placement decisions. Validating a test for specific purposes is an on-going process and is particularly important when a test undergoes significant revisions.

Construct validity is the extent to which a test represents the underlying traits or skills of interest (in psychometric terms, the “construct” itself). In order to develop a test with good construct validity, a theoretical basis for the construct is required. Since a construct is a hypothetical entity, it must be further operationalized by relating the construct to directly observable behaviors that serve as markers for the construct. On a test, answering questions correctly or incorrectly are the behaviors that are taken to reflect the underlying construct.

Evidence that is collected to evaluate construct validity can take multiple forms. One form of evidence is represented by the expert judgments of those who are knowledgeable about the

theory or theories regarding the underlying construct, and how this construct is measured by the test. Another form of evidence that can be used to support construct validity is to relate scores on the test in question with scores on other tests known or hypothesized to measure the same construct. If two or more tests are purported to measure the same construct, one would expect to find that the scores from these different tests would be highly correlated, and they should not be highly correlated with tests of similar formats that assess completely different constructs. A third type of construct validity evidence is if scores on the test are consistent with related psychological theories. For example, it would be expected that scores on a vocabulary test would increase as children completed more schooling. (Note that this result does not prove that the test is a good measure of a child's vocabulary, but rather is consistent with what is expected based on developmental psychology.) Different forms of construct validity evidence do not substitute for one another. Rather, the highest quality tests are those that have a strong foundation based on numerous studies that support the particular interpretation of scores on the test.

In the case of MAPP, since its format and structure are identical to the 2001 version of Academic Profile and similar to the 1990 version, we can assume that the validity evidence gathered for Academic Profile will likely be applicable for MAPP. For example, the findings from a study of the construct validity of the long form of Academic Profile would likely hold for MAPP (Marr, 1995). This study found that the higher scores of juniors and seniors could be explained almost entirely by their completion of more of the core curriculum, and that completion of advanced courses beyond the core curriculum had relatively little impact on Academic Profile scores. An earlier study (ETS, 1990) showed that Academic Profile scores increased as grade point average, class level and amount of core curriculum completed increased. In order to ascertain that these results are also true for MAPP, ETS plans to conduct additional

construct validity studies for MAPP in the near future. Studies of other aspects of validity, including content validity and predictive validity, will also be conducted as institutions continue to use MAPP and make their data available for ETS research studies.

## **References**

ETS. (1990). *The Academic Profile user's guide, part 1*. Princeton, NJ: ETS.

Marr, D. (1995). *Validity of the Academic Profile: An investigation of the construct validity of the long form of the Academic Profile*. Princeton, NJ: Educational Testing Service.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.), pp. 13-103. New York: Macmillan.