Harcourt ASSESSMENT

assessment report

Introduction to the *Miller Analogies Test*

Don Meagher, EdD
Senior Research Director
Post-Secondary Education

January 2006

Introduction to the *Miller Analogies Test*

Introduction

The *Miller Analogies Test* (MAT) is developed and administered by Harcourt Assessment, Inc. It has been used as a graduate school admission test for more than 70 years in university programs throughout the United States. The information presented in this paper is intended to familiarize graduate school admission directors, deans, and faculty with the MAT and to describe how the test is structured and administered. This paper also explains how the test results can be interpreted and used in admission decisions, provides information about the current norms, and discusses evidence regarding the reliability and validity of the MAT.

MAT is a high-level mental ability test requiring the solution of problems stated as analogies. Psychologists suggest that the analogy format represents an efficient and effective way to sample reasoning processes and to measure verbal comprehension and analytical intelligence (Gentner & Markham, 1997; Holyoak & Thagard, 1996; Lohman, 2004; Sternberg, 1977, 1985, 1988). Indeed, Robert Sternberg (1985) suggests that all of the information processing components involved in inductive reasoning are required to solve analogy problems: *encoding*, comprehending relevant information; *inference*, relating given concepts to other concepts; *mapping*, recognizing common rules shared by concepts; *application*, applying inferred rules to new concepts; *comparison*, choosing options that conform to ideals; *justification*, judging the reasonableness of choices; and *response*, demonstrating choices made by inductive reasoning. In a recent meta-analysis of the MAT, Kuncel, Hezlett, & Ones (2004) conclude that the analogical reasoning involved in the test involves all the principles of cognition, and that the MAT represents a valid predictor of performance in both academic and work settings.

The MAT Analogy

An analogy is a way of showing that two situations share a relational structure. Solving MAT analogies first involves recognizing a relationship between two given terms, a *base* analogue, and then looking for the same relationship between a third given term and one of four possible answer options, a *target* analogue. The correct answer must be selected by inferring the relationship between the two terms in the base analogue and *mapping* this relationship onto the one given term and the

option used to complete the target analogue. Mapping—correctly seeing the relationship between relationships—is the essence of solving the MAT analogy.

The Structure of an Analogy Item

A MAT analogy is a statement suggesting that two terms are related to each other in the same way that two other terms are related to each other. MAT analogy items are written as equations in the form "A : B :: C : D," which can be read as either "A is related to B in the same way that C is related to D" or as "A is related to C in the same way that B is related to D." A MAT analogy item is **never** written so that "A is related to D in the same way that B is related to C."

In a MAT item, one term in the analogy is missing and has been replaced with four options, only one of which correctly completes the analogy. Regardless of the order in which terms in an analogy are presented, there must be only one valid and logical relationship that exists between the two pairs of terms.

MAT analogies are intended to involve both content knowledge and analytical reasoning. Content knowledge is required to understand the meanings of the terms in the item, and analytical reasoning is necessary to recognize the analogical relationship between terms. For an analogy item to be effective in assessing analytical reasoning, it must involve recognizing relationships between terms in addition to an understanding of the meaning of the terms involved. To accomplish this, all the terms in the item must be familiar to the examinee. The difficulty of an analogy item is not only related to the difficulty of the terms involved, but also involves the subtlety of the relationship between the terms.

An Example of an Analogy Item

On the MAT, one term in each analogy item has been replaced with four options, only one of which is correct. The examinee is expected to select the option that creates a valid analogy. An example of a MAT analogy item is as follows:

> HOMOPHONE : (*a*. articulation, *b*. principle, *c*. significance, *d.* synonym) :: PRONUNCIATION : MEANING

The first step in solving this item is to decide which two of the three terms in the stem form a complete pair. In the example above, this could either be "Homophone is related to Pronunciation" (the first term is related to the third term) or "Pronunciation is related to Meaning" (the third term is related to the fourth term). Examinees are informed that an item will **never** be constructed so that the first term is related to the fourth term.

In this example, *Pronunciation* and *Meaning* are related in the sense that a pronunciation can have meaning, and *Homophone* and *Pronunciation* are related in

that homophones are words that are pronounced the same way. The only way to determine which of these relationships is the essential one in the analogy is to look at the four options.

In this item, the correct answer is "*d*. synonym" because a homophone is one of two or more words that have the same pronunciation, and a synonym is a word that has the same meaning as another word. Thus, the two terms in each pair are related in the sense that one term is a defining characteristic of the other. Each of the other three options has some relationship to one or more of the three given terms but does not form a valid analogy.

As this example illustrates, an effective MAT analogy involves an analytical reasoning process in the context of a specific content area—in this case, language.

What the MAT Measures

In studies of human intelligence and reasoning, researchers have found that verbal, quantitative, and figural analogies are among the best measures of verbal comprehension and analytical intelligence (Gentner & Markham, 1997; Holyoak & Thagard, 1996; Sternberg, 1977, 1985, 1988). Many cognitive psychologists also suggest that the cognitive process involved in creating analogies and recognizing conceptual similarities despite surface differences has practical benefits, including skill at solving problems, constructing explanations, and building arguments. The MAT has been designed to measure the same cognitive processes.

The MAT Content and Relationship Objectives

In addition to the cognitive processes involved in correctly solving MAT analogies, each item has been designed to also require background knowledge critical to the commencement of study in graduate school. To ensure that answering MAT items requires both the cognitive skills involved in recognizing analogical relationships and background knowledge appropriate for entering graduate students, MAT analogies contain content from the humanities, social sciences, natural sciences, mathematics, and language usage and vocabulary, as well as general knowledge commonly acquired through general reading and life experience.

The MAT test forms are composed of analogy items that reflect the content and relationship objectives listed in Figure 1.

MAT Content Objectives	
General (life experience, popular culture, work)	
Humanities (archaeology, art history, comparative religion, ethics, history, literature, modern and classical languages, philosophy, music, visual arts)	
Language (composition and rhetoric, grammar, word connotations, word meanings, word parts, word pronunciations and sounds)	
Mathematics (algebra, arithmetic, finance, geometry, numbers, probability, statistics)	
Natural Sciences (astronomy, biology, chemistry, earth science, ecology, environmental science, geology, physical geography, physics)	
Social Sciences (anthropology, civics, criminology, economics, education, geography, political science, psychology, public health, sociology)	
MAT Relationship Objectives	
Semantic (word meanings)	1. Similarity/Contrast (synonyms, definitions, similarities, antonyms, contrast, differences) 2. Intensity (size, degrees, semantic fractions/multiples) 3. Completion (parts of expressions, split phrases, words)
Classification (hierarchical relationship, classification, inclusion)	1. Category (member of class/class, class/member of class, subordination, superordination) 2. Membership (members of same class or category, coordination) 3. Whole-Part/Part-Whole
Association (ideas, predication, non-stationary set, processes)	1. Object/Characteristic (attribute, description, lacking quality, source, component, location, setting) 2. Order (sequence, reciprocal, by-product, transformation) 3. Agent/Object (cause/effect, creator/creation, function of, action taken by, purpose for, tool used by)
Non-Semantic (logical/ mathematical)	1. Equality (logical/mathematical equivalence, numerical fractions, multiples) 2. Negation (logical/mathematical negation) 3. Letter/Sound (letter patterns, sound patterns, rhymes, homophones, similar sounds)

Figure 1. MAT Analogy Objectives

The Development of the MAT

The MAT was originally developed by Dr. W. S. Miller at the University of Minnesota, where it was refined during the 1920s and 1930s and administered to all incoming graduate students in 1940. In 1947, The Psychological Corporation (now known as Harcourt Assessment) assumed publication of the test to manage the development and use of the MAT on a larger scale. At that time, a formal test construction procedure was implemented in conjunction with the author to ensure that future test forms would be constructed in a manner consistent with the original test. This procedure has ensured that all test forms currently in use retain the basic item types as previous forms, maintain a similar proportion of the content domains sampled, produce reliability estimates comparable to previous forms, and are directly comparable with current normative data.

The Structure of the MAT

The MAT is administered in 60 minutes, and each test form is composed of 120 analogy items, with 100 counting toward examinees' scores that are reported on personal Score Reports to the examinees and on Official Transcripts to institutions. The remaining 20 items are experimental and do not count toward examinees' scores. These items are being field-tested for possible use on future forms and are embedded within each test form so that examinees are unable to identify them.

Developing New Test Items

Harcourt Assessment has extensive experience in developing tests with a high degree of reliability and validity. New test items for the MAT are written by independent subject-matter experts, under the guidance of Harcourt Assessment professionals, to reflect the desired content and analogical relationships. Item writers are provided with standardized instructions to guide them in their writing efforts. These procedures have been designed to ensure the generation of high-quality test items that cover the range of difficulty and content necessary without requiring examinees to display extensive knowledge in any specific subject.

All new MAT test items are reviewed by independent subject-matter experts and by Harcourt Assessment measurement specialists for content appropriateness, and by editorial staff for style and format consistency. Each item is also reviewed to eliminate language, symbols, or content that may be considered offensive or inappropriate for sub-groups of the test-taking population. Only items that are judged as acceptable are then considered for field-testing on MAT test forms. After these experimental items are field-tested, the examinees' responses are analyzed for the statistical characteristics of each item. Both traditional test statistics and Item Response Theory (IRT) statistics are used to determine the difficulty and discriminating power of each item. Only the

items that have suitable psychometric properties and content are selected for the construction of new test forms.

Selecting Items for MAT Test Forms

Analogy items for MAT test forms are selected to be effective measures of both content knowledge and analytic reasoning. This is accomplished by selecting items according to very specific criteria:

- Solving analogy items requires subject matter and vocabulary knowledge that is reasonable to expect most American college students to have acquired through undergraduate general education and through general reading and experience.
- The correct answer to each analogy item involves recognizing a specific type of analogical relationship within the context of a specific content area.
- Each of the three incorrect answer options logically or semantically relates to at least one of the three terms in the stem, is consistent in structure with the other incorrect options and with the terms in the stem, and differs from the others in subtle ways that require fine semantic, logical, or relationship distinctions to determine that it is not the correct answer.
- Items must reflect appropriate psychometric properties and must represent a comparable range of statistical difficulty as items on the current MAT forms.

Test Administration

The MAT is administered through a network of more than 500 Controlled Testing Centers (CTCs) that have been established at colleges and universities throughout the United States and Canada (as well as at a few overseas sites) to serve their own students and candidates from the local area or region. When approved by Harcourt Assessment, each CTC is given the software necessary to administer computer-based versions of the MAT, and a complete supply of printed test materials to administer paper-and-pencil versions of the test. All testing materials must be maintained in a secure environment at all times and are inventoried annually by Harcourt Assessment.

Candidates make arrangements to take the MAT directly with the CTC most convenient to them. The CTC may administer either a computer-based or a paper-and-pencil version of the MAT individually or in groups. All CTC examiners are provided with a manual of directions to ensure adherence to the standardized administration procedures established for the test.

Reporting Test Results

After administering a paper-and-pencil version of the MAT, a CTC returns all examinee Answer Booklets to Harcourt Assessment for scoring and reporting. Answer Booklets are optically scanned and electronically scored. Following the administration of a computer-based version of the MAT, examinees' test data are immediately captured for processing by Harcourt Assessment. In both cases, the test results are stored in computer files from which personal Score Reports and Official Transcripts are generated. Results are mailed to examinees and to all designated score recipient institutions, typically within 15 business days of receipt by Harcourt Assessment. Examinees may request additional Official Transcripts whenever they wish, but scores that are more than five years old are not reported.

After many years of reporting MAT scores as raw scores and percentile ranks, a recent renorming of the test (2003) was followed by the introduction of scaled scores and the establishment of new percentile ranks based on the performance of recent MAT examinees. All personal Score Reports and Official Transcripts sent since October 2004 reflect the new norms and report examinees' MAT scores as a scaled score, as a percentile rank based on the total current norm group, and as a percentile rank based on the intended major indicated by the examinee. The new scaled scores range from 200 to 600 with a mean of approximately 400.

Personal Score Reports

Each examinee receives a personal Score Report after taking the MAT. This report is marked, "This is not an Official Transcript" to indicate that under no circumstances should an institution accept a personal Score Report as an Official Transcript.

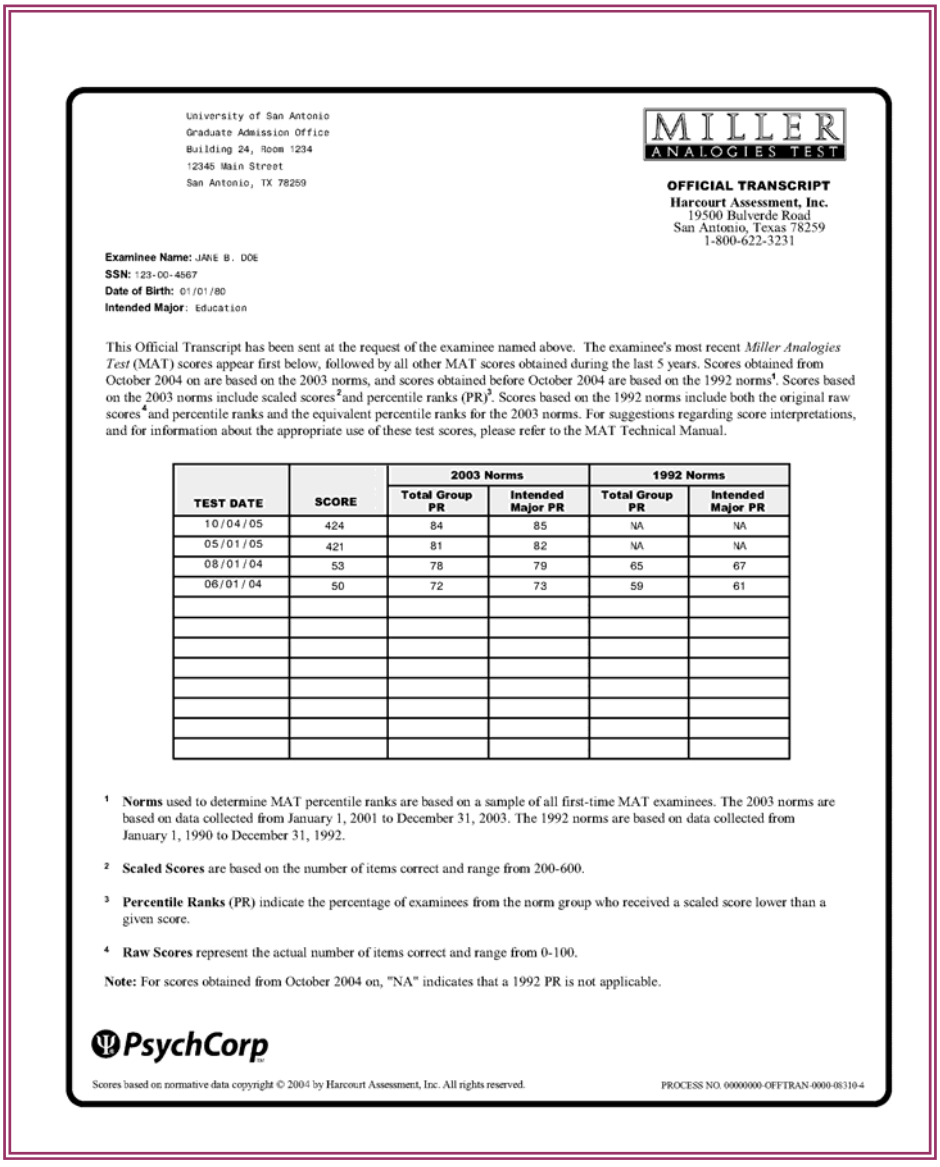
Examinees who take a computer-based version of the MAT immediately receive a Preliminary Score Report. The report clearly states that it is not an official score, that a personal Score Report will be mailed to the examinee, and that Official Transcripts will be mailed directly to the requested institutions.

Official Transcripts

Official Transcripts are sent to accredited institutions of higher education and approved fellowship or scholarship organizations that are requested by the examinee. To ensure score authenticity, institutions should accept only Official Transcripts sent by Harcourt Assessment. Scores listed on the Official Transcript indicate all MAT scores obtained during the five years previous to the date of the examinee's most recent attempt at the MAT. Under no circumstances are scores more than five years old reported to any institution or examinee. Harcourt Assessment does not recommend or establish a passing scaled score or percentile rank for the MAT.

As illustrated in Figure 2, the Official Transcript indicates the examinee's name, Social Security number, date of birth, and intended major (if any) as reported on the Answer Booklet at the time of testing. The Official Transcript also lists, in order by date, all the MAT scores obtained by the examinee during the previous five years.

Scores obtained since October 2004 are based on the current norms (2003) and are listed on the Official Transcript with scaled scores and percentile ranks. Scores obtained before October 2004 are based on the previous norms (1992) and are listed on the Official Transcript with the original raw scores and percentile ranks as well as with the equivalent percentile ranks based on the 2003 norms.



University of San Antonio
Graduate Admission Office
Building 24, Room 1234
12345 Main Street
San Antonio, TX 78259

MILLER ANALOGIES TEST

OFFICIAL TRANSCRIPT
Harcourt Assessment, Inc.
19500 Bulverde Road
San Antonio, Texas 78259
1-800-622-3231

Examinee Name: JANE B. DOE
SSN: 123-00-4567
Date of Birth: 01/01/80
Intended Major: Education

This Official Transcript has been sent at the request of the examinee named above. The examinee's most recent *Miller Analogies Test* (MAT) scores appear first below, followed by all other MAT scores obtained during the last 5 years. Scores obtained from October 2004 on are based on the 2003 norms, and scores obtained before October 2004 are based on the 1992 norms[1]. Scores based on the 2003 norms include scaled scores[2] and percentile ranks (PR)[3]. Scores based on the 1992 norms include both the original raw scores[4] and percentile ranks and the equivalent percentile ranks for the 2003 norms. For suggestions regarding score interpretations, and for information about the appropriate use of these test scores, please refer to the MAT Technical Manual.

TEST DATE	SCORE	2003 Norms		1992 Norms	
		Total Group PR	Intended Major PR	Total Group PR	Intended Major PR
10/04/05	424	84	85	NA	NA
05/01/05	421	81	82	NA	NA
08/01/04	53	78	79	65	67
06/01/04	50	72	73	59	61

[1] **Norms** used to determine MAT percentile ranks are based on a sample of all first-time MAT examinees. The 2003 norms are based on data collected from January 1, 2001 to December 31, 2003. The 1992 norms are based on data collected from January 1, 1990 to December 31, 1992.

[2] **Scaled Scores** are based on the number of items correct and range from 200-600.

[3] **Percentile Ranks (PR)** indicate the percentage of examinees from the norm group who received a scaled score lower than a given score.

[4] **Raw Scores** represent the actual number of items correct and range from 0-100.

Note: For scores obtained from October 2004 on, "NA" indicates that a 1992 PR is not applicable.

PsychCorp

PROCESS NO. 00000000-OFFTRAN-0000-08310-4

Figure 2. Sample MAT Official Transcript

Using MAT Scores in Graduate Admission Decisions

No single test score can possibly suggest all of the requisite knowledge and skills necessary for success in a graduate program. However, MAT scores can contribute significantly to the quality of the admissions process when used as part of an application package that consists of information from multiple sources about a candidate's academic achievement and potential. The MAT is intended to fairly and accurately reflect certain abilities acquired through individuals' educational experiences over an extended time. MAT scores should always be evaluated in the context of these experiences. Because some research suggests that MAT scores may not predict subsequent graduate school performance similarly for all groups of examinees (House & Keeley, 1995a; House & Keeley, 1995b; House & Keeley, 1996), MAT scores should be interpreted with consideration for each candidate's age, sex, ethnicity, economic background, linguistic background, educational attainments, and life experiences.

MAT Scores and Graduate Admissions

The MAT is typically used in admission decisions by graduate schools, often as an alternative to the *Graduate Record Examinations*® (GRE®). The current edition of the MAT Technical Manual contains complete information regarding the use and interpretation of MAT test scores for graduate school admissions. The MAT Technical Manual is available on request to university officials involved in graduate school admissions decisions.

It is the responsibility of each school or department to determine how MAT scores are used. The use of a cut score above which all applicants must score to be accepted into a program is not recommended without consideration for other application information. If a cut score is established, candidates' scores should always be considered in the context of appropriate measurement data for the test, such as the standard error of measurement and data regarding the evidence of predictive validity of the test. In addition, a cut score should not be set so high as to eliminate applicants who show other evidence of the ability to succeed, and not so low as to be attainable by chance (i.e., more than 25% correct, or a Total Group scaled score of a least 376).

Graduate schools using the MAT are encouraged to examine the relationship between entering students' MAT scores and their subsequent performance in the academic program. This locally obtained information will provide the best assistance in score interpretation and will most effectively enable a program to differentiate candidates who are likely to be successful from those who are not likely to be successful in that program. Under no circumstances should the MAT be used as an exit examination that would create a barrier to graduation for any individual who has already demonstrated success in a graduate course of study.

MAT Scaled Scores and Percentile Ranks

MAT scaled scores and percentile ranks reflect the general academic and analytic ability of the examinees in the normative sample. Normative scores, such as scaled scores, enable examiners to compare an individual's scores to his or her peers in the same group. Basing the MAT scores on the performance of a normative sample enables graduate schools to have a stable, unchanging criterion against which applicants can be evaluated from year to year.

MAT Scaled Scores

The MAT scaled scores represent equal units on a continuous scale, ranging from 200 to 600, with a mean of 400 and a standard deviation of 25. Because the construction of multiple test forms (unique collections of items) and formats (computer-based or paper-and-pencil) of the same test may result in slight variations in the level of difficulty, a given raw score from two different test forms or for the same form in two different formats may not always have precisely the same meaning. These differences are adjusted for in the scaled scores so the same scaled score has the same meaning regardless of the test form or format taken.

Scores for different MAT forms were equated by converting the raw scores (the number of items answered correctly in each subtest) to a common scale that is based on the performance of the current norm group. MAT norm group raw scores were converted to ability scores using IRT (the 1-parameter logistic model, also known as the Rasch model), ranging from approximately –5.0 to 5.0. The IRT ability scores were then used to derive the scaled scores. This equating method makes it possible to compare the MAT scaled scores of one test form to another and from one format to another.

MAT Percentile Ranks

The MAT percentile ranks were also derived from the IRT ability scores. The percentile ranks indicate the relative ranking of examinees in the norm group and range from 1 to 99. Although percentile ranks are useful for explaining an individual's performance relative to others, they have limitations. Percentile ranks do not have equal intervals. In a normal distribution of scores, percentile ranks tend to cluster around the 50th percentile. This affects scores in the average range most because a difference of one or two raw score points may produce a large change in percentile rank. Extreme scores are less affected, where a change of one or two raw score points typically does not produce a large change in percentile ranks.

The Current MAT Norm Group

Because the MAT measures academic content and analytical skills critical to success in graduate education, it is necessary to periodically establish new norms that accurately reflect the ability and performance of current graduate school applicants.

The current MAT scaled scores and percentile ranks are based on a group of more than 126,000 first-time MAT examinees who are U.S. citizens and took the test between January 1, 2001 and December 31, 2003. Data from the seven MAT forms used with the normative sample have been combined to produce the current norm references. The MAT Technical Manual contains detailed information about the current norm group.

Most of the percentile ranks were generally higher for the same raw scores for the 2001–2003 norm group than for the previous norm group (all first-time test-takers in 1990–1992). This reflects differences between the population taking the MAT in 2001–2003 and the population taking the test in 1990–1992. More recent examinees are generally answering fewer items correctly on the MAT than their peers did in 1990–1992.

Table 1 shows the number of examinees (*n*), the mean scaled scores (*M*), the standard deviations (*SD*), and the minimum and maximum, observed scores for the total norm group and for examinees in each of the reported fields of study for the 2001–2003 norm group.

Table 1. Distribution of Scaled Scores by Intended Field of Study and Total Group for the 2001–2003 Normative Sample

INTENDED FIELD OF STUDY & TOTAL GROUP	*n*	Scaled Score			
		M	*SD*	Min. Score	Max. Score
Business	8,280	396.5	25.3	237	511
Education	75,130	399.9	24.2	233	547
Humanities	2,967	413.7	26.2	326	547
Natural Sciences	7,614	402.8	22.8	331	529
Social Sciences	11,535	398.8	26.0	235	513
Other[1]	19,880	400.0	27.0	231	563
Undecided	676	399.4	27.0	320	506
Total Group	126,082	400.0	25.0	231	563

Note: [1] The "Other" group includes all majors that are not categorized under any other field of study.

Evidence of Reliability

The *reliability* of an assessment tool refers to the accuracy and precision of test scores and is a widely used indicator of the confidence that may be placed in those scores. The reliability of a test is expressed as a *reliability coefficient* that represents the consistency of scores that would be obtained if a test could be given an infinite number of times. Reliability coefficients can range from 0.00 to 1.00. A perfectly reliable test would have a reliability coefficient of 1.00, and a completely unreliable test would have a reliability coefficient of 0.00. Because the graduate school admissions process is so important, reliability coefficients of 0.90 or better are

typically expected from admissions tests. Reliability coefficients for the current MAT test forms range from 0.89 to 0.94.

Internal Consistency

The most commonly used formula for estimating the reliability of a test based on internal consistency is the Kuder-Richardson formula 20 (K-R 20). This method requires only a single administration of a test and requires information based on the number of items in the test, the standard deviation of the total score, and the proportion of examinees correctly answering each item. By examining the homogeneity of the questions within a test, K-R 20 yields reliability estimates of internal consistency called *reliability coefficients*.

Also related to the reliability of a test is the *standard error of measurement* (*SEM*)—an estimate of the possible amount of error present in a test score, or the amount that scores would probably vary if an examinee was tested repeatedly with the same test. Because repeated testing always results in some variation, no single test event ever measures an examinee's actual ability with complete accuracy. For this reason, a statistical allowance must be made for a test score to represent, with a reasonable degree of certainty, an estimate of an examinee's ability. The *SEM* is an amount that is added to and subtracted from an examinee's test score to determine a score range that represents a reasonable estimate of his or her actual ability. The lower the *SEM*, the closer an examinee's test scores are to his or her actual ability, and the greater the degree of certainty that the test scores are reliable.

Table 2 presents the reliability coefficients (K-R 20) and standard errors of measurement (*SEM*), along with mean scaled scores (*M*) and standard deviations (*SD*), for the MAT test forms currently in use.

Table 2. Reliability Coefficients (K-R 20), Standard Errors of Measurement, Mean Scaled Scores, and Standard Deviations for Current MAT Test Forms

	Current MAT Test Forms[1]							
	12	**18**	**21**	**22**	**23**	**24**	**25**	**26**
K-R 20	0.93	0.92	0.94	0.91	0.91	0.89	0.93	0.89
SEM	7.38	7.10	6.96	6.96	7.08	6.60	6.77	6.90
M	401.2	400.9	402.7	398.8	401.3	401.5	398.5	399.9
SD	27.9	25.1	28.4	23.2	23.6	19.9	25.6	20.8
n	3,844	16,224	8,304	21,992	597	632	568	650

Note: [1] Data for forms 12, 18, 21, and 22 from the 2001–2003 normative sample; data for forms 23, 24, 25, and 26 from more recent comparability studies (2004 and 2005).

As shown in Table 2, the reliability coefficients for the current MAT forms range from 0.89 to 0.94, and the *SEM* range from 6.60 to 7.38 scaled score points. The *SEM* were calculated at the 68% *confidence interval*—the range of certainty within which one

can have confidence that a score would occur upon repeated attempts by the same person. This means that at the 68% level of confidence, one can infer with about 68% certainty that the obtained score is plus or minus approximately 7 scaled score points from the examinee's theoretical true score. For the MAT test forms currently in use, the reliability coefficients are satisfactorily high, and the standard errors of measurement are sufficiently low.

Test-Retest

The test-retest method is another commonly used procedure to estimate test reliability. Two early independent research studies attempted to analyze the test-retest reliability of the MAT. Wallen and Campel (1967) compared test-retest MAT scores and found a mean score gain of 7 points on the second test. Dopplet (1971) found that the MAT had high test-retest reliability with reliability coefficients ranging from 0.82–0.89. Both studies concluded that the retest score gains were not significant.

In 2004 and 2005, Harcourt Assessment conducted a series of common-person comparability studies involving paper-and-pencil and computer-based test formats and newly developed test forms. A test-retest analysis of data collected for the most recent of these studies showed a reliability coefficient of 0.88.

Evidence of Validity

The evidence of the validity of a test may be defined as the degree to which the test actually measures what it is intended to measure. There are several types of evidence of the validity of a test. For the purposes of the MAT, evidence of construct, content, predictive, and consequential validity are discussed below.

Construct Validity

Construct validity is the degree to which the test measures a meaningful theoretical construct or characteristic. Many cognitive psychologists suggest that the ability to think analogically has practical benefits in activities such as problem solving, constructing explanations, and building arguments. According to Sternberg (1985), solving analogy problems involves all seven of the information processing components characteristic of inductive reasoning: *encoding* (comprehending relevant information to enable interpretation), *inference* (relating a given concept to another concept), *mapping* (recognizing a common rule shared by two concepts), *application* (applying a rule inferred from one set of concepts to another set of concepts), *comparison* (choosing an option that best conforms to an ideal), *justification* (determining the reasonableness of a choice relative to an ideal), and *response* (expressing a choice determined through inductive reasoning).

According to Lohman (2004), analogies represent "one of the most efficient item types" and allow "one to sample the efficacy of both past and present verbal reasoning processes across a much broader range of domains than could ever be represented in a necessarily smaller sample of reading passages" (p. 43). Lohman considers the recent trend moving away from using the analogy as an item type in standardized testing, and away from aptitude testing generally, to be a mistaken attempt to make admission testing more equitable. He argues that aptitude tests have an advantage over achievement tests in making predictions about success in new situations and "have an important role to play in admissions decisions, especially for minority students" (p. 41).

In a study of the effect of assessment type on the performance of minority examinees, Lohman (2004) compared the performance of White and minority students on a general achievement test (including items with content from various academic subject matter) and on a battery of cognitive ability tests (including various types of verbal, quantitative, and figural items, including analogies). Lohman found that for White students, it makes little difference which test is used with regard to performance in relation to pre-determined cut scores. However, he found that scores based on the cognitive ability tests resulted in consistent "increases over the achievement test in the percent of Black students who would be selected" if scores above the 70th percentile were a criterion for admission (p. 50).

In a meta-analysis of 127 studies, Kuncel, Hezlett, & Ones (2004) found that the MAT measures abilities that other cognitive ability instruments measure, and suggest that analogical reasoning involves all the principles of cognition. They also suggest that the MAT is an especially useful measure of general cognitive ability because it is composed of analogy items that require reasoning with vocabulary knowledge as well as knowledge of various domains. The authors conclude that a strong relationship exists between general cognitive ability and the acquisition of knowledge and skills (learning); that there are consistent positive relationships between general cognitive ability and academic and work performance; and that the MAT can be a valid predictor of performance in both settings.

Content Validity

Content validity is the degree of correspondence between the contents of the test and the logical and curricular domains intended to be measured. The MAT analogy items have been designed and constructed to measure knowledge, skills, and abilities considered necessary for success in graduate school. (See the sections in this paper, "The Development of the MAT" and "What the MAT Measures.")

Predictive Validity

Evidence of criterion-related or *predictive validity* shows how well MAT scores predict graduate school grades, professor ratings, degrees awarded, departmental

evaluations, or other indicators of subsequent success. Many studies conducted over the years have shown positive correlations between MAT scores and subsequent academic performance (Kuncel, Hezlett, & Ones, 2004).

The MAT and Other Predictors

Robertson and Hall (1964) found that using a combination of MAT scores, GRE scores, and undergraduate GPAs was the most promising predictor of faculty ratings, peer ratings, and comprehensive examination scores. In a nine-year study, Furst and Roelfs (1979) compared the predictive validity of the MAT to the GRE used in a doctoral program in education. Both the MAT and GRE scores showed moderate to low correlations with other criterion measures. In another study, DeCato (1982) suggested that neither MAT scores, GRE scores, nor GPA predicted performance in a specific course, but that these variables could be useful in assessing a general factor of scholastic ability. One study that examined the relationship between the MAT and the *Graduate Management Admission Test*® (GMAT®) found significant relationships between graduate GPA and both the MAT and GMAT (Graham, 1991).

Other studies have compared the MAT to other predictors in various specific majors. In a study by Littlepage, Bragg, and Rust (1978), the researchers used several variables to predict graduate school and professional performance in psychology, including GPA, MAT scores, the *English Cooperative Test*, GRE-Q (quantitative), faculty and employer ratings of students, and a survey administered to the graduate students concerning job duties, salary, job title, and undergraduate major. The researchers found that the MAT and an undergraduate major in psychology were significant predictors of graduate school performance, and that MAT scores, the undergraduate major, and faculty ratings were also significant predictors of professional performance. Tyron and Tyron (1986) indicated that both MAT and GRE scores could be used to predict a psychologist-trainee's ability to readily engage clients and that high engagers might be identified at admission to graduate school by using these variables. In another study that focused on a psychology program, Huber (1999) used MAT scores, GRE scores, and GPA to predict applicant's GPA in a doctoral program in clinical psychology. The results indicated that MAT scores were the most useful predictor of students' subsequent GPA in this program.

The MAT as Predictor for Various Groups

Research focusing specifically on the MAT suggests that MAT scores are useful in predicting both graduate course grades and the scientific contribution of students. In one early study, Platz, McClintock, and Katz (1959), considered GPA, doctoral examinations, faculty ratings, and MAT scores as predictors of graduate course grades and scientific contributions. The researchers found that MAT scores were the single best predictor of these outcomes. Focusing on programs in education,

Murray (1979) examined the variability of MAT scores within six departments of education and found significant differences within the departments. Students entering programs in educational psychology, secondary education, and instructional technology had significantly higher mean scores on the MAT than those in special education, elementary education, and educational administration programs. The researcher concluded that schools should carefully consider candidates in terms of the different areas of education to which they apply. Research by House and Keeley (1993) found that MAT scores were significantly correlated with graduate student performance. Other research findings have suggested that students in master's of education programs with higher MAT scores were more imaginative, intuitive, and more abstract in their thinking (Hughes, Costner, & Douzenis, 1988).

A Recent Meta-Analysis

A recent meta-analysis conducted by Kuncel, Hezlett, and Ones (2004) examined whether a test of general cognitive ability that was developed for use in educational settings is predictive of performances in both educational and occupational settings. The researchers defined general cognitive ability as "the underlying trait that leads to the well-documented positive intercorrelation observed between measures of cognitive behaviors" (Kuncel et al., p. 148).

Based on empirical evidence, the researchers concluded that a strong relationship exists between general cognitive ability and the acquisition of knowledge and skills. The researchers also suggest consistent positive relationships exist between general cognitive ability and both academic and work performance. Contrary to the view that academic tasks are different from real world tasks, the researchers suggest that the factors related to academic performance are similar to those of job performance because in both settings, performance is a direct function of acquired knowledge and skills. The researchers suggest that because the MAT assesses general cognitive ability, it can be a valid predictor of performance in both settings, even though it was originally developed for use in academic admissions.

The researchers examined 127 studies involving more than 20,000 participants and used statistical tools to correct for any sampling error, restriction of range, and measurement error in the various studies. This analysis found the MAT to be a valid predictor of several aspects of graduate student performance as well as measures of job performance, potential, and creativity.

They found correlations between the MAT and several indications of graduate school performance to be moderate to high. Correlations with first-year GPA were slightly higher than for overall graduate GPA, correlations with faculty ratings of graduate student performance were positive, and correlations between the MAT and comprehensive examination scores were the highest. Correlations with degree attainment and time to

degree attainment were found to be modestly positive, possibly moderated by motivational variables other than cognitive ability.

Table 3 shows predictive validity data for the academic criteria included in the meta-analysis of MAT research by Kuncel, Hezlett, and Ones (2004, p. 155) and from an earlier meta-analysis by the same researchers (2001, p. 169) of the GRE and entering undergraduate GPA.

Table 3. Mean Scaled Scores by Age Category

Criterion	MAT[a]			GRE[b]						UGPA[b]		
				Verbal			Quantitative					
	N	*k*	*p*	*N*	*k*	*p*	*N*	*k*	*p*	*N*	*k*	*p*
Graduate GPA	11,368	70	0.39	14,156	103	0.34	14,425	103	0.32	9,748	58	0.30
1st-year Graduate GPA	2,999	34	0.41	45,615	1,231	0.34	45,618	1,231	0.38	42,193	1,178	0.33
Degree Attainment[a]	3,963	20	0.21	6,304	32	0.18	6,304	32	0.20	6,315	33	0.12
Faculty Ratings	1,909	25	0.37	4,766	35	0.42	5,112	34	0.47	3,695	22	0.35
Comprehensive Exam Scores[cd]	987	10	0.58	1,198	11	0.44	1,194	11	0.26	592	6	0.12
Time to Complete[e]	1,700	5	0.35	160	3	0.28	160	3	–0.12	629	5	–0.08

Notes: MAT = *Miller Analogies Test,* GRE = *Graduate Record Examinations;* UGPA = undergraduate grade point average; *k* = number of studies; *p* = estimated true score validity for MAT, and estimated operational validity for GRE and UGPA (calculated by the researches to correct for sampling error, restriction of range, and measurement error in the data analyzed from the various studies). The six criteria presented in this table include only those with data available for all three predictors—MAT, GRE, and UGPA.

[a] From Kuncel, Hezlett, and Ones (2004, p. 155). [b] From Kuncel, Hezlett, and Ones (2001, p. 169). [c] Not corrected for criterion unreliability but with restricted range-restriction distribution (MAT); not corrected for range restriction (GRE and UGPA). [d] Not corrected for criterion unreliability. [e] GRE data only from the social sciences.

The researchers found that the MAT correlates highly with the Verbal section of the GRE and with other cognitive ability tests from educational and work settings. The researchers also suggest that selecting students or workers based on cognitive ability leads to the selection of individuals who are creative and have high potential.

Concerning transitional situations between graduate school and work, the researchers found the MAT to have a small, positive relationship with internship and practicum ratings, moderate validity for predicting counseling work sample performance, and a near zero correlation with student-teaching performance (again possibly due to other motivational variables unrelated to cognitive ability). The researchers also examined the relationship of the MAT to creativity ratings made by both faculty and work supervisors. They found the MAT to be a moderate predictor of ratings of overall potential, with the strongest correlation with counseling potential ratings. The researchers suggest that counseling performance was strongly

predicted and that the MAT was less strongly associated with educational administration performance. They conclude that the MAT is a generally valid predictor of work performance criteria.

The researchers state that "The abilities measured by the MAT predict more than just grades" but also predict "other academic criteria as well as a variety of professional and work performance criteria, including evaluations of creativity and potential" (Kuncel et al., p. 157).

Consequential Validity

Consequential validity refers to the effects that the assessment has on students, programs, and instruction. One common criticism of standardized tests is that they do not assess the key objectives of a given curriculum. While this criticism may be relevant to achievement tests that are intended to assess pervious learning in specific content areas, it is less so for a test such as the MAT that is more of an assessment of general cognitive aptitude. Nevertheless, admission tests such as the MAT do often create stressful situations for examinees who may feel that their future depends, in part, on the score they obtain. It is important to consider the degree to which the abilities measured with the MAT correspond to those valued by the program and utilized in the curriculum.

References

DeCato, C. M. (1982). Admissions criteria and later performance: Graduate Record Exam, Miller's Analogies, and GPA as predictors of professional competency. *Psychological Reports, 51*(3, Pt. 2), 1149–1150.

Dopplet, J. E. (1971). Differences between the Miller Analogies Test scores and people tested twice. *Educational and Psychological Measurement, 31*(3), 735–744.

Furst, E. J., & Roelfs, P. J. (1979). Validation of the Graduate Record Examinations and the Miller Analogies Test in a doctoral program in education. *Educational and Psychological Measurement, 39*(1), 147–151.

Gentner, D., & Markham, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45–56.

Graham, L. D. (1991). Predicting academic success of students in master of business administration program. *Educational and Psychological Measurement, 51*(3), 721–727.

Holyoak, K. J., & Thagard, P. (1996). *Mental leaps: analogy in creative thought.* Cambridge, MA: MIT Press.

House, J. D., & Keeley, E. J. (1996). Differential prediction of adult student performance from Miller Analogies Test scores. *Journal of Genetic Psychology, 157*(4), 501–503.

House, J. D., & Keeley, E. J. (1995a). Age bias in prediction of grade performance from Miller Analogies Test scores. *Journal of Genetic Psychology, 156*(2), 257–259.

House, J. D., & Keeley, E. J. (1995b). Gender bias in prediction of graduate grade performance from Miller Analogies Test scores. *Journal of Psychology, 129*(3), 353–355.

House, J. D., & Keeley, E. J. (1993). *Differential Prediction of Graduate Student Achievement from Miller Analogies Test Scores.* Paper presented at the annual meeting of the Illinois Association for Institutional Research, November 4–5.

Huber, L. L. (1999). A regression equation used as a criterion for admission to a graduate program in clinical psychology. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 60*(2-A).

Hughes, T. M., Costner, M. J., & Douzenis, C. (1988). Personality and motivation variables in tomorrow's teachers: A center of excellence. *Education, 180*(3), 393–403.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin,* Vol. 127, No. 1, 162–181.

Kuncel, N. A, Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1), 148–161.

Littlepage, G. E., Bragg, D. M., & Rust, J. O. (1978). Relations between academic performance, and professional performance. *Teaching of Psychology, 5*(1), 16–20.

Lohman, D. F. (2004). Aptitude for College: The Importance of Reasoning Tests for Minority Admissions. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 41–55). London: Falmer Press.

Murray, R. E. (1979). Variability in MAT results within the field of Education. *Psychological Reports, 45*(2), 665–666.

Platz, A., McClintock, C., & Katz, D. (1959). Undergraduate grades and the Miller Analogies Test as predictors of graduate success. *American Psychologist, 14,* 285–289.

Robertson, M., & Hall, E. (1964). Predicting success in graduate study. *Journal of General Psychology, 71*(2), 359–365.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ: Lawrence Erlbaum.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge: Cambridge University Press.

Sternberg, R. J. (1988). *The Triarchic Mind: A New Theory of Human Intelligence.* New York: Viking.

Tyron, G. S., & Tyron, W. W. (1986). Factors associated with clinical practicum trainees' engagements of clients in counseling. *Professional Psychology: Research & Practice, 17*(6), 586–589.

Wallen, N. E., & Campel, M. L. A. (1967). Vocabulary and non-verbal reasoning components on verbal analogies tests (Miller Analogies Test and Concept Mastery Test). *The Journal of Educational Research, 61,* 87–89.

800-211-8378