MAT

MILLER ANALOGIES TEST®

MAT Reliability and Validity

Effective: January 2021

Pearson

For inquiries or reordering:
800.622.3231
www.pearsonassessments.com/mat

NCS Pearson, Inc. 5601 Green Valley Drive Bloomington, MN 55437

Produced in the United States of America.

Table of Contents

Tables

Acknowledgments

This document was prepared by the following members of the Pearson Clinical Assessment group:

Donald G. Meagher, EdD
Senior Research Director

Natividad Ybarra III, PhD
Test Development Manager

Rachel P. Wegner, MA
Research Director

Tianshu Pan, PhD
Senior Psychometrician

Jeffrey R. Miller, MA
Digital Editor

Ariane Zamot, MEd
Digital Publishing Analyst

Introduction

The *Miller Analogies Test* (MAT) is a norm-referenced standardized test that is developed to measure the analytical skills and academic content critical to success in graduate education. To ensure the continuing relevance and usefulness of the MAT for assessing the prerequisite knowledge and skills of candidates for admission to graduate programs, the test content is periodically reviewed and revised.

This document provides information and research results related to the reliability and validity of the MAT. Information in this document includes reliability and validity data based on the current normative sample—MAT candidates who took the test for the first time between January 1, 2016 and December 31, 2019 ($N = 58{,}804$)—and a summary of research findings related to the validity of the test.

Additional publications are also available on the MAT website, including:

- **Understanding Analogies**—showing the usefulness of the analogy item format for measuring important cognitive skills

- **MAT Basics**—including information about MAT history, contents, structure, administration, and score reporting, as well as information useful in interpreting MAT scores

Also available only to qualified professionals, the printed *MAT Technical Manual* contains detailed data for the current normative sample, the current percentile ranks, and a compendium table that can be used to compare the previous percentile ranks with the current percentile ranks.

To request a copy of the *MAT Technical Manual*, or to offer suggestions regarding the MAT or about this or any other related publications, please contact MAT Customer Relations at MATScoring.Services@Pearson.com.

MAT Reliability

A test's reliability is the extent to which the test yields consistent results across multiple test administrations, for different methods of test delivery, or within a test form. This section will examine MAT reliability in terms of internal consistency and standard error of measurement, two statistics appropriate for use with a test like the MAT that candidates take in a single computer-based administration.

Internal Consistency

One common measure of a test's reliability is internal consistency, which is appropriate to use when candidates take a multiple-choice test in a single test administration. A commonly used formula for estimating the reliability of a test based on internal consistency is the Kuder-Richardson formula 20 (KR-20) that requires only a single administration of a test. KR-20 yields a reliability estimate of internal consistency (or reliability coefficient) by examining the homogeneity of the questions within a test based on the number of items in the test, the standard deviation of the total score, and the proportion of candidates correctly answering each item. KR-20 analysis assumes that the items on a test form vary in difficulty, and the resulting reliability coefficient represents an average of all possible split-half reliability estimates.

Reliability coefficients can range from 0.00 to 1.00, with a value of 1.00 indicating that the test is perfectly consistent. The reliability coefficients shown in Table 1 represent the internal consistency of all MAT test forms administered during the 2016–19 normative sample period for the total normative sample (Total Group) and for the six intended graduate major groups that are reported on Official Transcripts. All of the reliability coefficients shown in Table 1 are at least 0.90, and all are satisfactory according to commonly accepted criteria (Anastasi & Urbina, 1997; Streiner, 2003).

Table 1 **Reliability Coefficients (KR-20) and Standard Errors of Measurement (*SEM*s) for the MAT 2016–19 Normative Sample Scaled Scores**

Total group/intended field of study	*n*	Reliability	*SEM*
Total Group	58,791	.91	6.83
Business and Management	2,686	.90	7.44
Education	24,986	.91	6.77
Humanities and Arts	3,080	.92	6.99
Natural and Applied Sciences	7,306	.90	6.84
Social and Behavioral Sciences	16,707	.91	6.81
Other Fields	2,962	.93	6.80

Standard Error of Measurement (*SEM*)

The standard errors of measurement (*SEM*) shown in Table 1 also relate to the reliability of a test. The *SEM* represents an average discrepancy between a candidate's true score and the score the candidate actually obtains. Because every test is imperfect, no single test event ever measures a candidate's actual ability with perfect 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 a candidate's ability. The *SEM* is used with a candidate's test score to calculate a score range that includes a reasonable estimate of his or her true score. Thus, the smaller the *SEM*, the closer candidates' test scores are to their actual ability, and the greater degree of certainty that the test scores are reliable.

For example, if a MAT candidate obtains a scaled score of 400, the Total Group *SEM* of 6.83 shown in Table 1 suggests that one could be about 68% confident that the candidate would earn, upon repeated attempts, a score between 393 and 407 (i.e., 400 ± 6.83) and be 95% confident of a score between 387 and 413 (i.e., 400 ± 6.83 (1.96)).

MAT Validity

The validity of a test is defined as the degree to which the test actually measures what it is intended to measure. There are several types of test validity evidence. For the purposes of the MAT, evidence of construct, content, and predictive validity are examined in this section.

Construct Validity—the MAT Analogy

Construct validity is the degree to which the test measures a meaningful theoretical construct or characteristic. The construct validity of the MAT depends on the value of the analogy format in assessing abilities that are essential to success in graduate school. Evidence from research suggests that it does (Kuncel & Hezlett, 2007a, 2007b; Kuncel et al., 2004).

Studies that involve analogies typically structure analogy problems first with a base or source analogue that involves a specific type of relationship between two terms. Then in a target analogue that is composed of one given term and one missing term, the missing term must be selected from among options. The correct option 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 correct option used to complete the target analogue. Mapping—seeing the relationship between relationships—is the essence of an analogy (see also the "Understanding Analogies" document). It is this mapping ability that is significant with regard to human intelligence, which is believed to be dependent on the ability to ascertain the relationships between subjects (Lu et al., 2019). Analogies are also thought to be the driving force for the development of concepts that lead to human thought (Hofstadter and Sander, 2013). 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 et al., 2001; Gentner & Markham, 1997; Holyoak & Thagard, 1996; Kmiecik et al., 2019; Kollias & McClelland, 2013; Resing et al., 2017; Richland & Simms, 2015; 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 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:

1. Encoding—comprehending relevant information to enable interpretation
2. Inference—relating a given concept to another concept
3. Mapping—recognizing a common rule shared by two concepts
4. Application—applying a rule inferred from one set of concepts to another set of concepts
5. Comparison—choosing an option that best conforms to an ideal
6. Justification—determining the reasonableness of a choice relative to an ideal
7. Response—expressing a choice determined through inductive reasoning

In a meta-analysis of 127 studies involving more than 20,000 participants, Kuncel et al. (2004) found that the MAT measures abilities that other cognitive ability instruments measure, and suggested that analogical reasoning involves all the principles of cognition. They also suggest that the MAT is a 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, including the sciences, literature, arts, and history. 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

Another important type of validity is content validity—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. Such traits include the ability to analyze and interpret information, to think logically and critically, and to understand relationships across a broad spectrum of subjects (see the "MAT Basics" document).

To ensure that MAT test items involve the knowledge, skills, and abilities considered necessary for success in graduate school, all items have been written and reviewed by contracted individuals with content expertise according to detailed guidelines provided to them by Pearson. Prior to field-testing new test items, all items are reviewed by Pearson subject-matter experts and editorial staff for content appropriateness, style and format consistency, and gender and ethnic bias. Only items that are judged satisfactory through this process are considered for inclusion as items on MAT test forms.

Predictive Validity

An important way of evaluating the validity of a high-stakes standardized admission test is to examine its predictive validity, a type of criterion validity that addresses how well a test, like the MAT, predicts later criteria, such as grade point averages or program completion. Many studies conducted over the years have found the test to be a moderate to strong predictor of subsequent performance, with predictive validity statistics comparable to other standardized admission tests commonly used by graduate and professional schools (Kuncel et al., 2004; Kuncel & Hezlett, 2007a, 2007b).

Studies have compared the predictive validity of the MAT to other graduate school admission tests. Robertson and Hall (1964) found that a combination of MAT scores, *Graduate Record Examinations*® (GRE®) scores, and undergraduate grade point average (GPA) 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 criterion measures. In another study, DeCato (1982) suggested that neither MAT scores, GRE scores, nor previous GPAs 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, with correlations of 0.282 and 0.410, respectively (Graham, 1991).

In a comparison of meta-analyses of several standardized tests commonly used for graduate school admissions, Kuncel and Hezlett (2007a, 2007b) found nearly identical correlations (observed and corrected for measurement unreliability and restriction of range) between first-year graduate GPA and MAT scores (0.29 observed, 0.41 corrected), GRE General Test Total scores (0.27 observed, 0.41 corrected), and GMAT scores (0.27 observed, 0.41 corrected). They also found overall graduate GPA to

correlate slightly higher with MAT scores (0.27 observed, 0.39 corrected) than with either GRE Total scores (0.25 observed, 0.37 corrected) or GMAT scores (0.25 observed, 0.35 corrected).

Some studies have compared the MAT to other predictors in various specific majors. In an early study of graduate psychology majors, Platz et al. (1959) considered previous GPA, doctoral examinations, faculty ratings, and MAT scores as predictors of graduate course grades and scientific contributions. These authors found that MAT scores were the single best predictor of these outcomes. In a study by Littlepage et al. (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, faculty and employer ratings of students, 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 previous 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.

Research focusing on the MAT for use with programs in education suggests that MAT scores are useful in predicting graduate course grades. 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 MAT scores 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 of education programs with higher MAT scores were more imaginative, intuitive, and abstract in their thinking (Hughes et al., 1988). In two longitudinal studies involving candidates applying to a doctoral program in educational leadership that bases acceptance decisions on a compensatory model, the researchers found the MAT scores to be a stronger predictor of acceptance (Young, 2007; Young & Young, 2010) and graduation (Young, 2007) than undergraduate GPAs.

The publisher of the MAT has also conducted studies involving MAT scores and subsequent GPA. In one study (The Psychological Corporation, 1992) more than 50 graduate school departments provided information related to the predictive validity of the MAT. Predictive validity in this case was described as the statistical correlation between MAT test score and undergraduate GPA (the predictors) and the first-year graduate GPA. In a more recent study by Pearson (Meagher, 2008), nine graduate schools submitted data for students entering graduate programs in the fall of 2005, including undergraduate GPAs, previous graduate GPAs, MAT scaled scores, GRE scaled scores, demographic information, and GPAs for the first academic year in a graduate program. These data were also analyzed to determine the validity of the MAT and other entering variables for predicting subsequent GPAs using correlation analyses.

Table 2 shows the correlation finding for both the 1992 Psychological Corporation and 2008 Pearson studies. The 1992 study found correlation coefficients with first-year graduate GPA of 0.23 for MAT scores and 0.29 for previously-earned undergraduate GPA. The study completed in 2008 found that the variable with the highest correlation with first-year graduate GPA was previously-earned graduate GPA (0.30), followed by MAT scores (0.27), and GRE-Quantitative scores (0.27). The similar correlation

findings from these two studies suggest the consistency of both MAT scores and previously-earned grade point averages in predicting subsequent performance measured by first-year grade point averages. Whether based on the old raw scores obtained exclusively from paper-and-pencil administrations or on more recently-reported scaled scores obtained primarily from computer-based administrations, correlations between MAT scores and first-year graduate school performance have remained moderately positive.

Table 2 **Correlations (*r*) Between Predictor Variables and 1st Year Graduate GPA from the 1992 Psychological Corporation Study and the 2008 Pearson Study (* = significant at 0.01 level)**

		1st year graduate GPA
Predictor Variables	***n***	***r***
1992: MAT Raw Score	1,035	.23*
1992: Undergraduate GPA	938	.29*
2008: MAT Scaled Score	513	.27*
2008: GRE Verbal Scaled Score	437	.21*
2008: GRE Quantitative Scaled Score	432	.27*
2008: GRE Analytical Writing Score	326	.11
2008: Undergraduate GPA	639	.24*
2008: Previous Graduate GPA	118	.30*

The correlations found in both of these studies are also similar to findings in the research literature, particularly in comparison to correlations found in studies of the GRE and in meta-analyses of MAT and GRE research. Correlations between GRE General Test scores and graduate GPAs are typically found to be around 0.30 (Bridgeman, Burton, & Cline, 2009). In their meta-analysis of MAT studies, Kuncel et al. (2004) a found sample-size weighted average correlation of 0.29 between MAT scores and first-year graduate GPA, and an estimated true score validity (corrected for measurement unreliability and restriction of range) of 0.41 between MAT scores and first-year graduate GPA. In their meta-analysis of GRE research, the same researchers (Kuncel et al., 2001) found correlations between GRE scores and first-year graduate GPAs that were similar to those found in the MAT study (sample-size weighted average correlations of 0.24; estimated subtest operational validities ranging from 0.34 to 0.38), which suggests that MAT scores are at least comparable in predictive strength to GRE scores.

The Kuncel et al. (2004) meta-analysis represents the most extensive analysis of the predictive validity of the MAT conducted to date. In this study, the researchers found the MAT to be a valid predictor of several aspects of graduate student performance, job performance, potential, and creativity. They found that the MAT correlated highly with the GRE-Verbal section, with other cognitive ability tests from educational and work settings, and with the acquisition and demonstration of knowledge and skill. The authors also found that correlations between MAT scores and first-year GPA were slightly higher than for overall graduate GPA, that correlations with faculty ratings of graduate student performance were positive, and that correlations between MAT scores and comprehensive examination scores were the highest. Correlations with degree attainment and time to degree attainment were somewhat positive, possibly moderated by motivational variables other than cognitive ability.

These researchers also found the MAT to have a small, positive correlation with internship and practicum ratings, a moderate correlation with counseling work sample performance, and a near-zero correlation

with student-teaching performance, again possibly due to other motivational variables unrelated to cognitive ability. In addition, the authors examined the relationship of the MAT to faculty and work supervisors' ratings of creativity. They found the MAT scores to moderately correlate with ratings of overall potential, with the MAT most highly correlated with counseling potential ratings and less strongly correlated with educational administration performance. The authors concluded that the MAT is a valid measure of both general cognitive ability and verbal ability and that it is useful for predicting performance in both academic and work settings.

The correlations found in these and other studies are comparable to correlations typically found in similar types of analyses, where predictor variables are standardized test scaled scores with a large range of score points (e.g., MAT scaled score range of 200–600) and the outcome variable is a GPA (typically on a more restricted scale of 1–4). Regardless of the limitations inherent in such correlation analyses, consistently similar results found over many years suggest that the MAT continues to have value in predicting graduate program GPAs, particularly for the first year of graduate study, as well as other important academic and nonacademic performances.

Glossary

Construct Validity—The degree to which a meaningful theoretical construct or characteristic can be inferred from a test score.

Content Validity—Evidence for validity that relates to how adequately the content of a test represents a specified body of knowledge, and to how adequately subjects' responses represent knowledge of the content.

Correlation—A measure of the strength and direction of the relationship between two variables. (See Correlation Coefficient)

Correlation Coefficient (*r*)—A statistic ranging between –1 and 1 that indicates the degree and direction of relationship between two variables. The strength of the relationship is indicated by the values of the coefficients (with greater values indicating stronger relationships). The direction of the relationship is indicated by either a positive sign (+) representing a positive relationship in which variables tend to increase or decrease together, or a negative sign (–) representing an inverse relationship between variables.

Internal Consistency—A type of test score reliability indicating the degree of correlation among item scores from different parts of a test.

Kuder-Richardson Formula 20 (KR-20)—A statistic (ranging from 0 to 1) used to determine internal consistency reliability as an estimate of the average of all possible split-half coefficients, with a high value indicating a strong relationship among test items and a low value indicating a weak relationship. (See Correlation Coefficient)

Mean (*M*)—The average of a set of scores computed by adding all of the scores together and then dividing by the total number of scores.

Meta-Analysis—A method of research that analyzes the results of several independent studies by combining them to determine an overall effect or the degree of relationship between variables.

N-count (*N*)—The total number of individuals who make up a sample (e.g., the number of candidates who took a test).

Normative Sample/Norm Group—The group of individuals (sample) earning scores on a test whose score data are used to determine scaled scores and/or percentile ranks.

Norm-Referenced Standardized Test—A measurement in which an individual's scores are interpreted by comparing them to scores obtained by a defined group of individuals (a norm group or normative sample) whose score data are used to determine scaled scores and/or percentile ranks.

Percentile Rank (PR)—A whole number between 1 and 99 that represents the proportion of individuals from the normative sample who earned lower than a given score on a test.

Predictive Validity—Evidence for validity based on how accurately test data (e.g., admission test scores) are able to predict criterion measures obtained at some later time (e.g., a grade point average earned after admission).

Predictor Variable—A variable that is used to predict some subsequent outcome (e.g., a grade point average earned after admission).

Raw Score (RS)—The number of items answered correctly by a candidate on a test.

Reliability—An estimate of the dependability of test scores in terms of the degree of consistency between administrations.

Reliability Coefficient—A statistic (usually ranging from 0 to 1) that measures the degree to which test scores are free of measurement error. (See Standard Error of Measurement)

Scaled Score (SS)—A standardized test score on a specified common scale (e.g., 200–600) with a designated mean and standard deviation that are derived from a raw score (or an ability estimate). Scaled scores are especially useful for comparing performance of individuals or groups over time.

Standard Deviation *(SD)*—A measure of the variability of test scores in terms of how spread out scores are from the mean.

Standard Error of Measurement *(SEM)*—An estimate (based on group data) of the variation in scores earned on repeated administrations of the same test by the same individual, in the absence of practice effects or change in the underlying ability.

Validity—Validity refers to the extent to which evidence supports appropriate interpretations and inferences from test scores regarding characteristics of a person measured (e.g., knowledge or ability) or performances other than those measured (e.g., subsequent performance or achievement).

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.

Bridgeman, B., Burton, N., & Cline, F. (2009). A note on presenting what predictive validity numbers mean. *Applied Measurement in Education*, *22*, 109–119.

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.

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., Holyoak, K. J., & Kokinov, B. N. (2001). *The analogical mind: Perspectives from cognitive science*. MIT Press.

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 a master of business administration program. *Educational and Psychological Measurement*, *51*(3), 721–727.

Hofstadter, D., & Sander E. (2013) *Surfaces and essences: Analogy as the fuel and fire of thinking*. Basic Books.

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

House, J. D., & Keeley, E. J. (1993, November). Differential prediction of graduate student achievement from Miller analogies test scores. Paper presented at the annual meeting of the Illinois Association for Institutional Research, Oakbrook Terrace, Illinois.

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.

Kmiecik, M. J., Brisson, R. J., & Morrison, R. G. (2019). The time course of semantic and relational processing during verbal and analogical reasoning. *Brain and Cognition*, *129*, 25–34.

Kollias, P. & McClelland, J. L. (2013). Context, cortex, and associations: a connectionist developmental approach to verbal analogies. *Frontiers in Psychology*, *4*, 857.

Kuncel, N. R., & Hezlett, S. A. (2007a). Standardized tests predict graduate students' success. *Science*, *315*(5815), 1080–1081.

Kuncel, N. R., & Hezlett, S. A. (2007b). Supporting online material for standardized tests predict graduate students' success. [Electronic version]. *Science*, *315*(5815).

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*, *127*(1), 162–181.

Kuncel, N. R., 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.

Lu, H., Wu, Y. N., & Holyoak, K. J. (2019). Emergence of analogy from relation learning. *Proceedings of the National Academy of Sciences*, *117*(1), 4176–4181.

Meagher, D. (2008). *Miller Analogies Test: Predictive validity study general report* (Pearson White Paper). NCS Pearson.

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.

The Psychological Corporation. (1992). *Miller Analogies Test validity study*. Unpublished research study.

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

Resing, W. C. M., Bakker, M., Pronk, C. M. E., & Elliott, J. G. (2017). Progression paths in children's problem solving: The influence of dynamic testing, initial variability, and working memory. *Journal of Experimental Child Psychology*, *153*, 83–109.

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *WIREs Cognitive Science*, *6*(2), 177–192.

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

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

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence*. Viking.

Streiner, D. L. (2003). Starting at the Beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

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

Young, I. P. (2007). Predictive validity of grade point averages and of the Miller Analogies Test for admission to a doctoral program in educational leadership. *Educational Research Quarterly, 31*(2), 44–54.

Young, I. P., & Young, K. H. (2010). An assessment of academic predictors for admission decisions: Do applicants varying in national origin and in sex have a level playing field for a doctoral program in educational leadership? *Educational Research Quarterly*, *33*(3), 39–51.