Test of Mathematical Abilities – Third Edition (TOMA-3)

Virginia Brown, Mary Cronin, and Diane Bryant

Technical Characteristics

The Test of Mathematical Abilities, Third Edition (TOMA-3; Brown, Cronin, & Bryant, 2012) is an easily administered, norm-referenced, assessment tool used to identify, describe, and quantify mathematical deficits in school age children. The TOMA-3 can be administered either individually or in a group, taking approximately 90 minutes to complete. The test provides raw scores, age/grade equivalents, percentile ranks, scaled scores, and a Mathematical Ability Index. The TOMA-3 can be used by any professional with some formal training in standardized test administration.

TOMA-3 Subtests and Composite

The TOMA-3 has four core and one supplemental subtest. Their results can be combined to form an overall Mathematical Ability Index.

1. Mathematical Symbols and Concepts - Students answer a series of questions that relate to mathematical signs, symbols, words, or phrases. Each question has four possible answers (A, B, C, D).

2. Computation - Students solve a series of problems that increase in difficulty. They write their solution to the problems in the space provided in the Student Response Booklet.

3. Mathematics in Everyday Life - Students answer a series of questions that relate to the use of mathematics in everyday life. Each question has four possible answers (A, B, C, D).

4. Word Problems - Students solve a series of increasingly difficult word problems. They write their solution to the problems in the space provided in the Student Response Booklet.

5. Attitude Toward Math (Supplemental) - Students express their attitudes about mathematics instruction and their self-perceptions regarding their own abilities and achievement. For each statement the student marks one of four boxes (Yes, definitely!; Closer to Yes; Closer to No; No, definitely!). There are no right or wrong answers.

Mathematical Ability Index - Represents a broad range of mathematical abilities.

Demographics

The TOMA-3 was normed and stratified by age on a representative sample of 1,456 students ranging from 8 years 0 months to 18 years 11 months in age. These students resided in 21 states. The characteristics of the normative sample with regard to gender, region, ethnicity, Hispanic status, exceptionality status, household income, and parent educational attainment are comparable to those reported in *The Statistical Abstract of the United States* (U.S. Bureau of the Census, 2011).

Reliabilities

The study of a test's reliability centers on estimating the degree of error associated with its score. When error variance is investigated, results are usually reported in terms of a reliability coefficient, which is a specific use of the common correlation coefficient. In our investigation of the TOMA-3's reliability, we calculated three types of correlation coefficients: (a) coefficient alpha, (b) test-retest, and (c) scorer difference.

Coefficient Alpha

This type of internal consistency reliability demonstrates the extent to which test items correlate with one another. It estimates the amount of test error associated with content sampling and content heterogeneity. Coefficient alphas for the TOMA-3 were computed using Cronbach's (1951) coefficient alpha method.

Coefficient alphas for the TOMA-3 scores were calculated at eleven age intervals using data from the normative sample. The coefficients were averaged using the z-transformation technique. All of the averaged alphas for all five subtests round to or exceed .90, a most desirable level of reliability. The averaged coefficient for the Mathematical Ability Index is .96, a value indicating nearly perfect reliability. The *standard error of measurement (SEM)* corresponding to these averaged alphas are 1 for the subtests and 3 for the composite.

Test-Retest

The test-retest method examines the extent to which an individual's test performance is consistent over time and is used to estimate time sampling error in a test. This approach involves administering the test and re-administering it a week or more later. The degree of similarity between the two test scores indicates the amount of stability possessed by the test.

We investigated this type of reliability using a sample of 51 individuals ranging in age from 8 years to 17 years, from Texas. Fifty-five percent were males; 73% were identified as White, 10% as Black/African American, 10% as Asian Pacific Islander, 7% as Two or More, and 25% were identified as Hispanic. Ten percent reported a disability.

The TOMA-3 was administered twice to the sample; the average intervening time was 14 days. After testing was completed, the standard scores were correlated at each testing. The resulting coefficients, with a single exception, exceed .80. The magnitude of these coefficients is large enough to support the idea that the test's scores contain little time sampling error.

Scorer Differences

Reliability among scores of objective tests is understandably high. In such instances, unreliable scoring is usually the result of clerical errors or improper application of standard scoring criteria on the part of the examiner. Scorer error can be reduced considerably by availability of clear administration procedures, detailed guidelines governing scoring, and opportunities to practice scoring. The correlation between the scorers is a relational index of agreement.

In the case of the TOMA-3, two members of the PRO-ED staff who were familiar with the test's scoring procedures independently scored 50 complete TOMA-3 protocols drawn at random from the total sample. The subjects represented a broad range of ability and ranged from age 9 to 17. Twenty-three were males, and 27 were females. The results of the two scorings were then correlated. The resulting coefficients which all exceed .90 in magnitude provide strong evidence supporting the test's scorer reliability.

Validity

This paper discusses three types of evidence for the relationship between TOMA-3 and mathematical ability: (a) review of correlation coefficients showing the relationship between TOMA-3 and these comprehension measures, (b) comparison of the means of TOMA-3 and these measures, and (c) results of a series of binary classification and ROC Area Under the Curve (ROC/AUC) analyses using TOMA-3 to predict mathematical ability.

Correlations with Mathematical Ability Measures

This study examined data collected during the norming process. The subjects of the criterion mathematical ability measures were 87 students ranging in age from 9 to 16 years from Alabama, Arkansas, Illinois, New Mexico, Pennsylvania, or Vermont. Forty-three percent were males; 75% were identified as White, 13% as Black/African American, 7% as American Indian/Eskimo/Aleut, 3% as Asian Pacific Islander, 2% as Two or More, and 66% were identified as Hispanic. Six percent of the sample reported a disability. This study examined the specific relationship of the TOMA-3 to the following mathematical ability measures:

- *Comprehensive Mathematical Abilities Test* (CMAT; Hresko, Schlieve, Herron, Swain & Sherbenou, 2003).
- *Iowa Algebra Aptitude Test - Fifth Edition* (Schoen & Ansley, 2007).

The correlations between TOMA-3 and mathematical measures answer a theoretical question: Does TOMA-3 measure mathematical abilities of school age children as well as other standardized measures? The coefficients have been attenuated for any lack of reliability in the criterion tests (but not in TOMA-3) and corrected to account for any range effects that might artificially depress or inflate the coefficients. Both corrected and uncorrected coefficients were calculated.

The coefficients representing the core subtests range from .50 (Large) to .74 (Very Large). The coefficients corresponding to the TOMA-3 Mathematical Ability Index are .92 (Nearly Perfect) for CMAT, .83 (Very Large) for IAAT-5, and .88 (Very Large) for the composite.

Comparisons of the TOMA-3 and Mathematical Ability Test Means and Standard Deviations

When two tests are highly correlated, they are likely to be measuring the same or a similar ability. This does not necessarily mean, however, that the tests yield the same results. For example, one test may consistently score higher than another test even though they correlate well with each other. The validity of both tests is supported when the two tests produce similar means as well as correlate highly with each other.

The standard score means, standard deviations, and comparative information for TOMA-3 and the mathematical ability measures were calculated. The differences between the means of the two tests were analyzed using the *t*-test and effect size correlation methods.

The difference between the means of the TOMA-3 and those of the mathematical ability measures were not statistically significant. These findings provide added support for the claim that TOMA-3 is a valid measure of mathematical ability. The findings support the idea that for all practical purposes, regardless of samples' characteristics or the criterion test administered, the standard scores that result from giving the TOMA-3 will be similar to those obtained from giving the criterion tests.

Binary Classification and Area Under the ROC Curve (ROC/AUC) Analyses

Binary classification analysis involves the computation of a test's sensitivity and specificity indexes. In the current context, the *sensitivity index* reflects the ability of a test to correctly identify individuals who exhibit a mathematical learning disability—the most important attribute of a screening test. The *specificity index* reflects the ability of a test to correctly identify individuals who do not have a mathematical learning disability.

The results for sensitivity and specificity are reported as portions (i.e., percentages). The size of the proportions necessary to be considered acceptable will vary depending on the purpose of the analysis (e.g., when screening for cancer, a relatively high number of false positives is tolerable in order to assure that the number of true positives identified is high). Authorities vary regarding how large a test's sensitivity and specificity indexes should be. Wood, Flowers, Meyer, and Hill (2002) and the Committee on Disabilities of the American Academy of Pediatrics (2001) recommend that the sensitivity and specificity indexes should be at least .70. Jansky (1978), Gredler (2000), and Kingslake (1983) prefer .75 for both indexes. Carran and Scott (1992) recommend a more rigorous standard of .80 or higher. Jenkins and others (Jenkins, 2003; Jenkins, Hudson & Johnson, 2007; Johnson, Jenkins, Petscher, and Catts, 2009) recommend that sensitivities should be high—perhaps as high as .90—and that specificity levels should be relatively high as well.

The area under the ROC curve (ROC/AUC) "is a measure of the overall performance of a diagnostic test and is interpreted as the average value of sensitivity for all possible values of specificity" (Park, Goo, & Jo, 2004, p.13). ROC/AUC values range from 0 (representing zero predictive ability) to 1 (representing perfect predictive ability). Zhou, Obuchowski, and Obuchowski (2002) recommend that screening measures designed to distinguish between students with and without mathematical deficiencies should have AUC values that are close to 1. Compton, D. Fuchs, L. Fuchs, and Bryant (2006) suggest that ROC/AUCs of .90 and above are considered excellent; .80-.89 are good; .70-.79 are fair; and .69 or below are poor.

If the TOMA-3 is a reliable and valid measure of mathematical abilities, the test results should differentiate between individuals who have a known diagnosis of mathematical learning disability from those who do not. To test this hypothesis, using the demographic information collected at the time of testing, individuals in the TOMA-3 normative sample were allocated into two groups, those with a diagnosis of mathematical learning disability (N = 60) and those without (N = 1,396). If the TOMA-3 can be used to identify individuals who have mathematical learning disabilities, examinees who score in the bottom 25% on the TOMA-3 should also have a diagnosis of mathematical learning disability. Therefore, a 2x2 frequency matrix was created with TOMA-3 scores classified as either (a) below the 25th percentile or (b) at or above the 25th percentile, and as associated with students who either (a) have or (b) have not been diagnosed mathematically learning disabled. Table 1 reports the results of this binary classification indexes and ROC/AUCs.

All of the values reported in the table meet or exceed the high standards recommended by authorities mentioned earlier in this section including low numbers of false positives and thereby provide strong and conclusive evidence of the TOMA-3's ability to identify students with a mathematical learning disability.

Discussion

While the validation of a measure is always ongoing, the data presented here indicate that the TOMA-3 is a promising new measure of mathematical abilities. It's ease of administration and ability to screen students are especially useful in identifying students who are significantly behind their age mates in mathematical knowledge and ability, determine the magnitude (below average, poor, or very poor) of any mathematical problems, or used in research investigating mathematical issues. Additional evidence of construct validity is reported in the TOMA-3 Examiner's Manual.

References

Brown, V. L., Cronin, M. E., & Bryant, D. P. (2012). Test of mathematical abilities (3rd ed.). Austin, TX: PRO-ED.

Carran, D. T., & Scott, K. G. (1992). Risk assessment in preschool children: Research implications for the early detection of educational handicaps. *Topics in Early Childhood Special Education*, *12*, 196-211.

Committee on Children with Disabilities, American Academy of Pediatrics. (2001). Developmental surveillance and screening of infants and young children. *Pediatrics, 108,* 192–196.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention. *Journal of Educational Psychology, 98*:2, 394–409.

Cronbach, L. K. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

Gredler, G. R. (2000). Early childhood screening for developmental and educational problems. In B.A. Braken (Eds.), *The psychoeducational assessment of preschool children* (pp. 399-411). Boston: Allyn & Bacon.

Hresko, W. P., Schlieve, P. L., Herron, S. R., Swain, C., & Sherbenou, R. J. (2003). *Comprehensive mathematical abilities test.* Austin, TX: PRO-ED.

Jansky, J. J. (1978). A critical review of some developmental and predictor precursors of reading disabilities. In A. L. Benton & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge* (pp. 412-516). New York: Oxford University Press.

Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students.* Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium, Kansas City, MO. Retrieved April 3, 2006, from http://www.nrcld.org/symposium2003/jenkins/index.

Jenkins, J. R., Hudson, R. G., & Johnson, E. S. (2007). Screening for service in an RTI framework: Candidate measures. *School Psychology Review, 35*, 582-599.

Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009) How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice, 24*(4), 174-185.

Kingslake, B. J. (1983). The predictive (in)accuracy of on-entry to school screening procedures when used to anticipate learning difficulties. *British Journal of Special Education, 10,* 24-26.

Park, S. H., Goo, J. M., & Jo, C-H. (2004). Receiver operating characteristic (ROC) curve: Practical review for radiologists. *Korean Journal of Radiology, 5*(1), 11–18.

Schoen, H., & Ansley, T. (2007). *Comprehensive Iowa algebra aptitude test* (5th ed.). Rolling Meadows, IL: Riverside Publishing.

U. S. Bureau of the Census (2011). *The statistical abstract of the United States*. Washington, DC: Author.

Wood, F., Flowers, L., Meyer, M., & Hill, D. (2002, November). *How to evaluate and compare screening tests: Principles of science and good sense.* Paper presented at the meeting of the International Dyslexia Association, Atlanta.

Zhou, X. H., Obuchowski, N. A., & Obuchowski, D. M. (2002). *Statistical methods in diagnostic medicine.* New York: Wiley.

Table 1 A listing of Classification Indices Demonstrating the Ability of the TOMA-3 Mathematical Ability Index to Predict Mathematical Learning Disability (Decimals Omitted)

TOMA-3 Cut Score	%ile Rank	n	ROC/AUC	Sensitivity	Specificity	True Positives	False Positives	True Negatives	False Negatives	Classification Accuracy
90	25	1,456	87	95	79	57	290	1,106	3	80

Note. ROC/AUC = Receiver Operating Characteristic /Area Under Curve.