**Test of Mathematical Abilities—2**

The Test of Mathematical Abilities–2 (V. Brown, Cronin, & McEntire, 1994) is a norm-referenced test intended for use with students between the ages of 8–0 and 18–11. The test differs from other math tests in that it goes beyond computation and solving of story problems to measure three other aspects of math now considered to be crucial: (1) students' attitudes toward math; (2) their understanding of the language of mathematics, as represented in the vocabulary of instruction; and (3) their familiarity with math terms and concepts used in everyday life. The authors indicate that the test is designed to assist school personnel in making decisions about curriculum placement, assessing the extent to which students have mastered specific math skills, and making normative comparisons of student performance in math. In addition, the test is designed to answer these questions (V. Brown et al., 1994, p. 1):

- What is the student's expressed attitude toward math?
- What is the student's general math vocabulary level?
- How knowledgeable is the student about the functional use of math facts and concepts?
- How do the student's attitudes, vocabulary, and general math information compare with the basic skills shown in the areas of computation and story problems?
- Do the student's math vocabulary and level of general math information differ from those of his or her peers?
Most of this test is identical to the first edition of TOMA. A few items were added to the beginning and end of each of four subtests. The authors indicate that they performed an item analysis on the original TOMA items and retained what they called "good" items. The authors do not indicate the number of items retained but do say that "TOMA-2 contains most of the items found in the original TOMA and additional items added to the top and bottom of each subtest" (V. Brown et al., 1994, p. 33).

Subtests

The 120 test items (13 more than in the original TOMA) are grouped into the following subtests:

Vocabulary  The student must write definitions of each of 25 mathematical terms, such as volume, ratio, and ellipse.

Computation  Students solve 25 problems in their answer booklets. The problems sample basic math skills, as well as advanced fractions, percentages, money, and complex math problems.

General Information  Students respond orally or in writing to questions about math as used in everyday situations. For example, "Why can a canceled check be your receipt?"
**Story Problems** Students solve 25 word problems in their answer booklets.

**Attitude Toward Mathematics** This is a 15-item supplemental test in which students respond to questions such as, "It's fun to work math problems" and "My friends like math more than I do," using a 4-point scale: "yes, definitely," "closer to yes," "closer to no," and "no, definitely."

**Scores**

Raw scores on the TOMA-2 can be converted to age scores, standard scores (with a mean of 10 and a standard deviation of 3), grade equivalents, percentiles, and math quotients.

**Norms**

The norms of TOMA-2 are a combination of TOMA norms and new norms. Remember that the only new items for TOMA-2 are a few at the beginning and end of four subtests. An undetermined number of items were dropped from the original TOMA. If students in the original norm sample reached a ceiling without taking higher-level items, their scores were included in the norms for TOMA-2. The performances of 336 students were used.
In addition, two procedures were used to develop the TOMA-2 norm sample. First, testing sites were identified in four major census districts. Normative data were collected by one of the authors or by a specially trained person in these sites. The authors say that the sites were chosen because the demographic characteristics of the students in the school matched closely those in the region. No data on sample selection are provided. The second method was to go to Pro-Ed customer files and pick people who had purchased TOMA and would be willing to test students in their own school districts. Each willing respondent tested 10 to 50 students. The standardization sample for TOMA-2 is not a stratified sample.

There were 2,082 students in the norm group. Tables in the manual indicate that their characteristics resemble closely the characteristics of students in the general population. No cross-tabulation information is presented, so we do not know whether most of the males came from the East, African American students from the West, and so forth. We believe that it is critical to indicate the makeup of the norm sample beyond one-variable descriptions.

Reliability

The authors report information on internal-consistency and test–retest reliability. Internal consistency was computed using data from the standardization sample. All coefficients exceed the desirable .80. Test–retest reliability is reported for 198 students in New Orleans. The students ranged from 9 through 14 years of age. Given that 72 percent were African American, whereas 14 percent of the norm sample was African American, the
sample was not representative. The test–retest reliabilities range from .70 (Attitude Toward Math) to .85 (Story Problems). Information about the reliability of this test is too limited to be useful. More studies need to be completed on larger and more representative samples.

Validity

A major improvement in this second edition of TOMA is seen in the rationale and explanation the authors provide for item selection. This was missing, for the most part, in the original TOMA. The authors claim that the test has good content validity because items are similar to those in other math tests. Yet, they do not report the relationship between items on TOMA-2 and the content of mathematics curricula.

In establishing the criterion-related validity of this test, the authors report studies on both the original TOMA and the TOMA-2. They argue that the tests are "almost identical." To establish comparability, they report correlations in performance on the two scales. Nearly all correlations are above .98. The results of two very limited studies are then reported. The first is a study of the relationship between TOMA performance and performance on now out-of-print versions of the KeyMath Diagnostic Test, Peabody Individual Achievement Test, and Wide Range Achievement Test. The study was completed on only 38 children, all of whom had learning disabilities and ranged in age from 9 through 14. The results are essentially meaningless as evidence for the validity of TOMA-2.
The second study of criterion-related validity was also completed on the original TOMA. The scores of an unspecified (age, grade, gender, ethnicity, SES?) sample of 290 students attending school in Illinois and Louisiana were correlated with their performance on the math subtests of the Science Research Associates achievement test. Correlations were moderate, but we do not know what they mean.

Evidence for construct validity is provided. The authors demonstrate that scores increase with age and that performance on the subtests is moderately intercorrelated. They demonstrate that the test correlates moderately with out-of-print versions of the Wechsler Intelligence Scale for Children and the Slosson Intelligence Test. There is no description of the group on whom this study was conducted, and the study is not referenced. As further evidence of construct validity, the authors again report the results of the study of 39 students with learning disabilities. They show that the students earned lower-than-average scores.

In our opinion, the validity studies conducted on TOMA-2 do not provide evidence of the test's validity. The studies are on very limited samples, which are not described; when samples are described, the age range is limited to older students; and when correlations with other measures are reported, the other measures are out-of-print editions of currently used tests. There is no evidence for the validity of the one unique subtest, Attitude Toward Math.

Summary
The TOMA-2 is designed to measure skill development in math and, in addition, attitude toward math. The authors contend that the test is virtually unchanged from its original version. In standardizing this measure, the authors did not use a stratified-normative-sampling approach. Thus, we do not know the extent to which the norms are representative. Evidence for reliability is limited to evidence of internal consistency. There is no evidence of test–retest reliability. Evidence of validity is limited to content validity. The remaining reliability and validity studies had major design problems. The SDMT4 appears to be a better diagnostic mathematic test.