The sixth edition of the Metropolitan Readiness Tests (Nurss & McGauvran, 1995) is intended to assess beginning reading, story comprehension, and quantitative concepts, three skill areas believed to be "essential for prereading and premathematics learning in the early school years" (p. 11). There are two levels of the test. Level I is individually administered and intended to assess skills needed before and during kindergarten. Level 2 is group administered and intended to assess skills needed from mid-kindergarten through beginning first grade.

Three subtests form the Beginning Reading composite. Story Comprehension and Quantitative Concepts and Reasoning are each assessed by one subtest named for the area. The subtests are briefly described here.

**Visual Discrimination** (Level I) This is a Beginning Reading subtest that assesses skill in matching individual letters, letter sequences, and words.

**Beginning Consonants** (Levels I and II) This is a Beginning Reading subtest that assesses skill in discriminating initial phonemes.

**Sound–Letter Correspondence** (Levels I and II) This is a Beginning Reading subtest that assesses skill in identifying letters that correspond to sounds.
Aural Cloze (Level II) This is a Beginning Reading subtest that assesses skill in matching the initial sound of a picture that is contextually appropriate for a sentence read by the examiner.

Story Comprehension (Levels I and II) This subtest assesses understanding of the vocabulary and concepts of a story to which subjects listen.

Quantitative Concepts and Reasoning (Levels I and II) This subtest assesses basic mathematical concepts (such as number–numeral and part–whole relationships) and operations (for example, addition).

Scores and Norms

Raw scores for Visual Discrimination, Beginning Consonants, and Sound–Letter Correspondence are summed to form the Beginning Reading composite. Story Comprehension and Beginning Reading composite are summed to form the Prereading composite score. All raw scores can be summed to form a Total Test composite, although there is no provision on the record forms for this score.

Raw scores and composites can be converted to several norm-referenced scores: percentiles, stanines, scaled scores that are indecipherable,1 and normal-curve equivalents that appear to be based on percentile ranks in a normal curve. In addition,

1 The authors explain that the “scaled score system links together common Skill Areas and the Prereading Composite across the two levels” (p.15). As in earlier editions, the authors do not provide means and standard deviations for the scaled scores and do not indicate how the two levels are linked.
there are two kinds of "content-referenced" scores. The first content-referenced score is the raw score. No rationale is provided for considering these scores content-referenced. The second type of content-referenced score, a performance rating, is a 3-point scale that indicates that a student currently has "learned enough of the skills . . . to be judged proficient" (a + rating), "is in the process of learning the skills" (a \check\ rating), or "needs instruction in the skills" (a 2 rating).

Apparently school districts were stratified and sampled to represent the U.S. population in terms of geographic region, socioeconomic status, urban/suburban/rural residence, ethnicity, and public–private school attendance. No further explanation of the sampling plan is provided, and no data are provided to demonstrate that the selected districts are indeed representative. Rather, the authors present tables to show that the children are generally representative of children in the United States; the actual norms are based on weighted samples.

Reliability

Internal consistencies (KR-20) for subtests and the Total Test composite for Level 1 were calculated for two prekindergarten samples (tested at midyear or spring) and three kindergarten samples (tested at fall, midyear, or spring); for Level 2, they were calculated for two kindergarten samples (midyear and spring) and for first-graders in the fall. Five of the eight internal-consistency estimates for the Total Test composite equaled or exceeded .90; 2 of the 40 internal consistencies of subtests equaled or exceeded .90, and 25 were less than .80. Test–retest reliability was estimated using 124 undescribed
students who were retested with Level 2 after an unspecified period. Only the stability of the Total Test composite was greater than .90; two subtests had stabilities less than .80. Thus, only the reliability of the composite is occasionally high enough for making important decisions for individual students.

Validity

Evidence for the validity of the MRT6 is inadequate. No evidence of content validity is presented; no evidence of construct validity is presented. No evidence of predictive validity is presented for Level I. The only validity evidence presented is for Level II and indicates that the MRT6 correlates moderately with the Metropolitan Achievement Tests, Seventh Edition, and the Stanford Achievement Test, Ninth Edition. Most troubling is the authors' failure to present any data whatsoever to support their interpretations of "content-referenced" scores.

Summary

The sixth edition of the MRT is the latest version of a test originally published in 1933. The technical qualities of the test are marginal. The reliability of the total composite is usually sufficient for making important decisions for individual students; the other scores are usually not reliable enough for that purpose. The norming procedures are poorly described. Validity evidence is largely absent.