**Detroit Tests of Learning Aptitude–4**

The Detroit Tests of Learning Aptitude–4 (Hammill, 1998) are the most recent version of Baker and Leland's test, originally published in 1935. To Hammill's credit, the 1985, 1991, and current revisions have systematically strengthened weaker areas of the test and risen to the increasingly rigorous standards of educators and psychologists. This edition of the test has more attractive stimulus materials, better-described norms, more reliability information, and additional validity information. The Picture Fragments subtest has also been dropped.

The test is described as a measure of developed abilities, a concept first used by Anastasi (1980). She argues that the considerable confusion created by using the terms *aptitude, intelligence, and achievement* could be reduced by substituting for each the word *abilities*. Hammill (1991) adopts Anastasi's notion of developed abilities and a fundamental assumption that goes along with the term. Anastasi (1988, p. 413) states that all ability tests—whether they be designed as general intelligence tests, multiple aptitude batteries, special aptitude tests, or achievement tests—measure the level of development attained by the individual in one or more abilities. No test reveals how or why the individual reached that level.

There are four principal uses for the test: "(a) to determine strengths and weaknesses among developed mental abilities, (b) to identify children and youths who are significantly below their peers in important abilities, (c) to make predictions about
future performance, and (d) to serve as a measurement device in research studies investigating aptitude, intelligence, and cognitive behavior" (Hammill, 1998, p. 24). The DTLA-4 consists of ten subtests that measure different but interrelated developed abilities in individuals ranging in age from 6–0 to 17–11.1 The DTLA-4 takes between 50 minutes and 2 hours to administer.

Subtests

**Word Opposites** A stimulus word is read aloud, and the student is asked to state a word that means the opposite of the stimulus word.

Design Sequences A student is shown a card with a sequence of designs for five seconds. The card is then removed, and the student uses cubes to reproduce the designs from memory.

Sentence Imitation The examiner reads a sentence without inflection, and the student must repeat the sentence.

Reversed Letters The examiner says a series of letter names at the rate of one letter per second, and the subject then writes each letter in the series in reversed order.

**Story Construction** The student is shown pictures and is asked to make up and tell stories in response to the pictures.
Design Reproduction  Geometric forms are presented for specified time intervals and then removed. The student must draw the forms from memory.

Basic Information  The student must answer specific factual questions that assess knowledge of everyday situations, rather than knowledge acquired in school.

Symbolic Relations  The student is shown a design and then must select from among six possible responses the pattern that completes the design.

Word Sequences  The student is required to repeat a series of unrelated and isolated words read by the examiner.

Story Sequences  The examinee is shown a series of cartoonlike pictures and must put these into sequence in order to depict a story. The student indicates the order by putting numbered chips under the pictures.
An examiner administers all items of the Design Sequences, Reversed Letters, Story Construction, and Story Sequences subtests. On the remaining subtests, examiners start testing at a specified point and continue until a ceiling is reached. On all tests with ceilings except Design Reproduction, the ceiling is that point at which the student has failed five consecutive items. On Design Reproduction, the ceiling is reached when the student has received a score of zero on three consecutive drawings.

Subtests can be combined into 6 different composites. All subtests are combined to form the General Mental Ability Composite. The Optimal Composite is composed of the four subtests on which a student earned the highest scores; these subtests, of course, vary from person to person. Subtests can be combined into 14 other composites. Three pairs of DTLA-4 composites contrast verbal and nonverbal, attention-enhanced and attention-reduced, and motor-enhanced and motor-reduced abilities. The remaining 8 composites are intended to represent the theoretical aspects of intelligence formulated by other theorists. Specifically, there are separate composites intended to reflect Cattell's crystallized and fluid intelligences, Das's simultaneous and successive processing abilities, Jensen's associative and cognitive levels of intelligence, and Wechsler's verbal and performance abilities.

Scores and Norms

Raw scores on each subtest can be converted to age equivalents, percentiles, and standard scores with a mean of 10 and a standard deviation of 3. Composite scores can be converted to percentiles and quotients with a mean of 100 and a standard deviation of 15.
DTLA-4 norms combine the individuals from two groups. The first group consists of the 922 students from the DTLA-3’s norms who were tested in 1989–1990. The second group consists of 428 students tested in 1996–1997. Students in the 1996–1997 sample were located using a sampling plan similar to the plan used to identify part of the 1989–1990 sample. Four sites (one in each geographic region of the United States) were selected, and classrooms within each area were identified. Thus, students came from intact classrooms. Students with disabilities are included in the norms to the extent that they were enrolled in these classes. The remainder of the 1989–1990 sample came from individuals across the United States who agreed to participate by testing 10–30 students. The number and location of these participants are not reported. The overall sample closely resembles the U.S. population in 1996.

While the description of the normative sample is substantially better than in previous editions, it remains less than complete. For example, geographic area, gender, race, ethnicity, urban/rural residence, and parental income resemble the U.S. population for four age groups (6–8, 9–11, 12–14, and 15–17). No data are presented to show that the sample is representative at each age. An appropriate sampling plan might provide the basis for a reader's inferring that the DTLA-4 samples at each age are representative because the samples for age groups are representative. However, the description of the sampling plan is too incomplete to make such an inference with much confidence. For example, it is not possible to tell if the "primary standardization sites" were restricted to the communities named or if they included the surrounding suburban and rural areas. It is also unclear how one primary standardization site (for example, South Wales, NY, or New Orleans, LA) is representative of an entire geographic region.
Reliability

The DTLA-4 manual contains information about internal consistency, stability, interscorer agreement, and differences between subtests and between composites. Internal consistency for each subtest at each age was estimated from the performance of the normative sample. The 120 coefficient alphas (10 subtests at 12 ages) range from .71 (Story Sequences at age 6 and Basic Information at age 7) to .97 (Design Sequences at age 10). Of 120 alphas, 9 (7.5%) are less than .80 and 41 (34.2%) equal or exceed .90. The internal consistency of all composites, with one exception, equals or exceeds .90. (The exception is when a 6-year-old's optimal composite uses the four most unreliable subtests; in that case, the optimal composite's internal consistency is .87.) Internal consistency was also examined separately for specific subgroups within the normative samples (males \( n = 687 \) and females \( n = 663 \); Euro-Americans \( n = 976 \), African Americans \( n = 198 \), Asian Americans \( n = 27 \), and Hispanic Americans \( n = 133 \); students with learning disabilities \( n = 63 \) and students with mental retardation \( n = 37 \)). A similar pattern of coefficients was found. With two exceptions, about half of the subtest reliabilities were in the .80s and half were in the .90s; the reliability of every composite equaled or exceeded .90. Although these coefficients are somewhat higher than the age group coefficients, they were calculated across ages.

To investigate stability, 96 students (from grades 3 to 12) from Austin, TX, were tested twice within one week. The students are described in the manual as multicultural and of low socioeconomic status. Test–retest coefficients are presented for all subtests for
three groups: 24 children in grades 1 through 3 (although the manual indicates elsewhere that there were no students in first or second grade), 36 students in grades 4 through 6, and 36 students in grades 7 through 12. Of the stability estimates, 10 were in the .70s, 11 were in the .80s, and 9 were in the .90s. No estimates of the composites for the three age groups are presented; presumably, they are higher. Stability estimates for the pooled groups are also presented; it is unclear why these estimates are substantially higher.

To investigate the interscorer reliability, two individuals independently scored 30 completed protocols (that is, answer forms) and converted the raw scores to standard scores. These standard scores were then correlated. Interscorer agreement for all subtests and composites equaled or exceeded .95. However, as described in the DTLA-4 manual, the examiners apparently did not evaluate the students' responses, but instead dealt with protocols that had already been scored. Thus, the coefficients seem more likely to reflect the degree to which basal and ceiling rules are applied, correct responses summed, and derived scores found for the subtotals and totals.

Finally, the author presents the critical values needed to evaluate the reliability of a difference between any two subtests or between any two composites. However, these values are based on the average reliability (across ages), not on the actual age of the student whose scores are being compared.

Validity

The DTLA-4 manual contains considerable information about various aspects of the test's validity. Because the DTLA-3 and DTLA-4 are quite similar, much of the evidence for
the present edition is actually from the previous edition. Rather than separate the studies by edition, we refer to the test as the DTLA when evidence from the two editions is combined.

Evidence for content validity is of two types. First, the author justifies the inclusion of the content of each subtest with a paragraph or two in which he shows that other tests of intelligence use similar types of items and/or provides a brief rationale for the content. Second, the author also shows how the items and subtests on the DTLA-4 cover the aspects or elements of intelligence that are most often incorporated into the various theories or descriptions of intelligence.

Evidence for criterion-related validity rests in part on the validity of the previous edition and in part on new research. Several studies (with sample sizes ranging from 25 to 50) correlated the DTLA and various measures of intelligence (the Kaufman Assessment Battery for Children, the Scholastic Aptitude Scale, the Peabody Picture Vocabulary Test–Revised, the Wechsler Intelligence Scale for Children–III, and the revised Woodcock–Johnson Psychoeducational Battery). The correlations between the General Mental Ability Quotient of the DTLA-4 and these measures varied from .55 to .91. Moreover, the pattern of correlations also tends to support the theoretical composites. For example, the WISC-III verbal IQ correlates more highly with the DTLA's verbal scale than with the nonverbal scale, while the WISC-III performance IQ correlates more highly with the DTLA's nonverbal scale than with its verbal scale.

A variety of evidence for construct validity is presented. First, the DTLA-4 subtests appear to measure the same constructs for African Americans, Hispanics, and females. Moreover, the means for males and females and for students from various ethnic
groups are remarkably similar. Second, subtest scores increase with age, as would be expected. Third, students identified as having mental retardation by independent criteria earn substantially lower scores; students identified as having learning disabilities by independent criteria earn scores between those earned by students with mental retardation and students who are not identified as having a disability.

The results of several studies indicate that the DTLA predicts school achievement as measured by the Wide Range Achievement Test–Revised, the Diagnostic Achievement Test–2, the Diagnostic Achievement Battery for Adolescents–2, and the Diagnostic Achievement Battery–2. The DTLA also predicts teacher ratings on the Comprehensive Scales of Student Abilities.

Finally, several factor-analytic investigations were conducted. Two factors appear to underlie the entire scale: verbal ability and nonverbal ability. Confirmatory factor analyses indicate that the other composite scores are tenable.

Summary

The Detroit Tests of Learning Aptitude–4 are intended to measure developed abilities. The test contains 10 subtests that can be formed into 16 composites. The sampling plan for the development of the norms is incompletely described. However, the resulting sample does approximate the U.S. population in terms of geographic area, race, gender, parental income, and urban/rural residence. The internal consistency of composites at all ages is excellent. Stability data are incompletely reported and weaker. Ample evidence
suggests that the DTLA-4 provides an unbiased measure of general intelligence as well as other ways of looking at intellectual functioning.