

Northwest Achievement Level Tests

The Northwest Achievement Level Tests (Northwest Evaluation Association, 1999) are group-administered tests designed to measure student progress in reading, language, math, and science achievement. The tests are modified to meet the curricular goals of a particular school or district; however, all items are selected from the same item bank. Brief locator tests assist in determining the specific achievement level test to be administered to each student. In this way, every student is administered a test that corresponds to his or her individual achievement level, regardless of current grade level. There are 16 levels for math, 8 for reading, and 7 for language. The test is intended for students in grades 2 through 10. The test is not timed, although administration time is estimated to be 60 minutes per content area. The authors indicate the following uses of the test: (1) to measure the achievement growth of individual students, classrooms, and grade levels; (2) to enhance instructional planning at the individual and classroom levels; and (3) to evaluate districts. Computer-adaptive tests that measure similar content are also available from the Northwest Evaluation Association (Measures of Academic Progress); however, the focus of this review is on the paper-and-pencil version of the test.

Subtests

For each subtest, students silently read and respond to approximately 50 multiple-choice items. Although specific tests may vary based on district goals, tests typically contain the following subtests and subtest components.

Reading This subtest measures achievement in the broad areas of word meaning, literal comprehension, interpretive comprehension (making inferences, summarizing,

synthesizing information, etc.), and evaluative comprehension (distinguishing fact from opinion, detecting bias, understanding persuasion, etc.).

Language This subtest measures achievement in the broad areas of writing process, composition structure, grammar/usage, punctuation, and capitalization.

Math This subtest measures achievement in the areas of numeration systems, operations/computation, equations/numerals, geometry, measurement, problem solving, statistics/probability, and applications.

Science Concepts This subtest measures achievement in the areas of science concepts (cause–effect, energy–matter, force, interaction, quantification, system, etc.) and science processes (classifying, defining operationally, hypothesizing, measuring, questioning, etc.).

General Science This subtest measures achievement in the areas of life science, earth/space science, and physical science.

Norms

The normative sample includes students from participating Northwest Evaluation Association member districts who contributed data from fall 1996 to fall 1998 test administrations. Approximately 545,200 students from 105 school districts in 18 states are represented in this sample. Science test normative information is based on a smaller data set. The manual provides the following demographic characteristics of the sample: ethnicity, grade, and state of residence. Ethnicity information and state of residence information are disaggregated by grade level. Only ethnicity information is compared to characteristics of the national population and bears only slight resemblance to the

population. No specific information on gender, racial, socioeconomic status, or community type characteristics of the norm group is provided.

Scores

Scores are based on the Rasch Unit (RIT) scale; these scores range from 120 to 320. All levels of the test are incorporated into this single scale of measurement to allow for measuring individual growth over multiple years. Third-grade scores typically range from 150 to 190; high school scores typically range from 240 to 300. Percentile ranks are available by grade level; however, these scores allow comparison to a group of same-grade students that does not necessarily represent characteristics of the national population. Growth scores (fall to spring and spring to spring) can be compared to the average growth of this comparison group.

Reliability

Marginal reliability and test–retest reliability estimates are provided in the manual for each grade level and content area (with the exception of science). These may vary slightly from district to district because the tests are individually designed. Marginal reliabilities were determined based on the 1996 normative sample, and all exceed .90 for each content area (reading, language, and math) at each grade level with the exception of grade 9 language scores (.89). Second- and tenth-grade marginal reliabilities are not provided. Test–retest reliability correlation coefficients are also displayed, and are based on correlating 1998 normative sample scores across fall and spring test administrations. These coefficients range from .70 (grade 2 mathematics) to .93 (grade 7 and 8 mathematics) and tend to be higher at the more advanced grade levels. Tenth-grade test–retest reliability information was provided only for the Math subtest.

Validity

Various types of validity information are provided in the manual. Evidence of validity based on test content. Content validity is addressed through a description of the item-development process. Items are written by classroom teachers and reviewed for racial/ethnic bias by teachers of various backgrounds. The items are then field-tested in order to evaluate their measurement characteristics, and attach them appropriately to the measurement scale. Validity based on relations with external measures is Criterion-related validity is evidenced by correlations of the NALT with other achievement tests. These include the Stanford Achievement Test–Ninth Edition (SAT9) and various statewide assessments. Correlations range from .78 to .88 with the SAT9 by grade level and content area. Average growth scores of students in the normative sample indicate that scores increase over the course of receiving instruction, suggesting some degree of evidence for validity based on internal structure.construct validity.

Summary

The Northwest Achievement Level Tests are designed to measure student progress in reading, language, math, and science achievement according to local curricular goals. The test is administered at each student's achievement level, as determined by a brief locator test. RIT scores are provided on a unitary scale to allow for measurement of student progress across years. Demographic information of the norm sample is not representative of the nation; therefore, the reporting of NALT percentile ranks should be avoided. Limited reliability and validity information is provided in the manual.