

Achieve is a criterion-reference test which meets the generally accepted standards of validity and reliability set forth in "Standards for Educational and Psychological Testing" (1999), published by The American Research Association. The test is multiple-choice in nature and covers areas in Reading/Language Arts and Mathematics for grades K-12. The test was developed by a group of educational experts and designed to meet state curriculum standards. In selecting test items, consideration was given to test validity and reliability. Validity refers to how well a test measures what it is designed to measure (Mason, 1989). According to Van Dalen (1979) an investigator may check one or more types of validity (content, criterion-related validity, and construct validity) in order to determine the test validity. Van Dalen further asserts that content validity is most widely used in achievement tests and mostly concerned with the degree to which a test measures an intended content area.

To ensure content validity, Achieve considered the amount of testing time and types of test items based on compatible standardized test used in the state. The group of educational experts who wrote the test items examined both the state curriculum adapted by the Department of Education and released items from previous used standardized test that the state has used for assessing student achievement. Test items are then written based on these considerations because the questions were appropriate for the target



population, clearly worded, and contained appropriate vocabulary. Each item is carefully designed to align with a state standard as dictated by K-12 grade Reading/Language Arts and Mathematics curriculum. In developing the test, a list of objectives on each unit was compiled to determine subject validity for these test items. A matrix was constructed linking objectives and test questions.

To determine an index as it relates to criterion-related validity, the test was administered to a group of students for comparison with a previous administered achievement test that had been adapted by the state. A statistical test was performed on the test scores of the students. The statistical analyses determined the predictive validity coefficient range for the Reading/Language Art tests (Words/Phrases, Main Idea/Purpose, Comparison, and Reference/Research) as 0.819 to 0.878.

A second important characteristic of a test is reliability, the dependability and the consistency, or trustworthiness of the score. Reliability is the degree to which a test consistently measures what it proposed to measures. That is if one used a test several times with the same sample population under the same treatment conditions, and the test was reliable, then one would get approximately the same score each time. The reliability of the test was determined from the application of one of the Kuder-Richardson formulas, which are based on the proportion of persons correctly answering each item and the standard-deviation of the scores. The results were analyzed by statistical software package and the reliability coefficient was 0.81 for the Reading/language Arts and 0.79 for the Mathematic Assessment test.