

## Section 2. Validity

Validity is one of the most important attributes of assessment quality. Validity refers to the degree to which evidence supports a proposed interpretation or use of a set of scores, and it is one of the most fundamental considerations in developing and evaluating tests (AERA, APA, & NCME, 1999). Validity is not based on a single study or type of study; but involves an ongoing process of gathering evidence supporting the interpretation or use of the resulting test scores. The process begins with the test design and continues throughout the entire assessment process, including design, content specifications, item development, psychometric quality, and inferences made from the results.

The development of test content for each MDHSA was overseen by a content expert who had a depth of knowledge and teaching experience related to the course in which the MDHSA was to be administered. Appropriate content leads who had similar qualifications reviewed the test development work of these individuals.

The test development process provided numerous opportunities for the client to review test content and make changes to ensure that the items were valid measures of the knowledge and skills of Maryland students according to course standards. Every item that was created was referenced to a particular instructional standard (i.e., goal, expectation, or indicator). During the internal ETS development process the specific reference was confirmed or changed to reflect changes to the item. When the item went to a committee of Maryland educators for a content review, the members of the committee made independent judgments about the match of the item content to the standard it was intended to measure, and evaluated the appropriateness for the age of students being tested. These judgments were tabulated and reviewed by the content experts who used the information to decide which items would advance to the field test stage of development.

The constructs measured by each MDHSA are described in detail in the Maryland high school curriculum standards (Core Learning Goals). All ETS content staff working on item development has been trained in the Core Learning Goals. The test blueprint documents presented in Section 1 (see Tables 1.2 to 1.5) were created in collaboration with committees of Maryland educators and were derived from the Maryland goals, expectations, and indicators. The Learning Goals can be found on the MSDE website at <http://www.mdk12.org>.

Although all eligible students participated in the MDHSA and information regarding student performance was provided to students, parents, teachers and other stakeholders, scores had no consequences for individual students. Algebra and English scores were used for AYP as a component of the Maryland No Child Left Behind (NCLB) Accountability program. Information on the interpretation of scores was provided to students, parents, schools and other stakeholders via the MSDE website.

In addition to the validation documentation gathered and maintained by MSDE, empirical information in support of the MDHSA follows:

- Section 3 provides detailed information concerning the scores that were reported for the MDHSA and the cut-scores for each content area.
- Section 4 provides demographic information for the population of students who were administered the MDHSA. Summary statistics at the test level are reported for the student population and for subgroups. In addition, score reliability analyses and measures of decision accuracy and consistency are provided for the student population.
- Section 5 includes documentation regarding the field test analyses. Descriptions of classical item analyses, differential item functioning, item response theory calibration and scaling are included. In addition, summary tables of item p-values and item-total correlations distributions are provided.
- Section 6 describes the development of the new English test. The section covers the procedures used for selecting operational items, calibration and linking, summary statistics of student achievement, decision accuracy and consistency, IRT analyses, and factor analyses.