Appendix D
Technical Aspects of Statewide Assessments

Technical Aspects

The following material is intended to assist those called upon to explain to others issues such as referencing, reliability, and validity. We have provided a general overview in order to promote better understanding of ISTEP+ and of assessment in general. For further information see also *Appendix H* for the 2009 *ISTEP+ Reliability and Validity Report* submitted by CTB/McGraw-Hill.

Test Score Referencing: A Question of Meaning

A test score, in and of itself, has very little meaning. For a score to derive meaning, it must be referenced to an outside criterion agreed to and understood by those who interpret the results. Depending on the criterion selected, we can determine how a student performed in comparison to other students, or in comparison to a standard that defines success or mastery.

Criterion-Referencing: Setting a Goal for Performance

Criterion-referenced tests seek to yield information about what a particular child can or cannot do within a given subject area. Once an area of learning has been defined, test items can be written and selected in such a way that scores take on meaning in terms of knowledge of a particular subject matter. If we identify the score that corresponds to mastery of the subject area, we can tell how a student's performance compares to mastery.

Vertical Scaling (ISTEP+)

Vertical scaling is the psychometric step needed to link all statewide assessments so that longitudinal growth may be measured from grade level to grade level. In Fall 2002, the ISTEP+ assessment program initiated a vertical scaling project that resulted in grade-to-grade alignment of test scores. To construct the scale, test items drawn from each grade level's academic standards were developed and administered to representative samples of students in Grades 3-10 from across Indiana. Student responses were ordered sequentially, one level at a time, to create a vertical scale that connects all on-level tests for those grades. A new vertical scale was created during the spring 2015 ISTEP+ administration based on college- and career-ready 2014 Indiana Academic Standards in English/Language Arts and Mathematics for students in grades 3-8. The scale will be expanded in 2015-16 when grade 10 is added.

Referencing ISTEP+

Among other things, ISTEP+ is intended to assist teachers in recognizing a student's strengths and areas of need relative to the *Indiana Academic Standards* to better drive instruction within the classroom. This purpose requires information about individual students and their respective mastery of English/language arts, mathematics, science, and social studies. ISTEP+ also provides information that is criterion-referenced relative to the *Indiana Academic Standards*.

Reliability: A Question of Accuracy

Reliability reflects how accurately a test measures the content it was intended to measure. In order for a test to be useful, then, it must be reliable. A test administered over and over to the same student (assuming no learning has taken place between test administrations) should produce similar (although

not identical) scores. Reliability also can be demonstrated in other ways, such as by giving an examinee two halves of a test—the items of each selected at random—and then comparing the two scores. To ensure effectiveness, ISTEP+ items are subjected to various tests of reliability.

Content Validity

Content validity, unlike its psychometric counterpart, questions whether a test measures what is important and valid in the curricula. ISTEP+ has evolved continuously, since its inception in 1987, to measure more accurately that which is important and valued by Indiana educators and curriculum experts. All criterion-referenced reports indicate critical concepts and knowledge identified by the Indiana State Board of Education as *Indiana Academic Standards*. The Department's content and assessment staff, working with educators from around the state, continually strive for the best match possible between that which should be taught and that which is tested.