

CHAPTER 17: VALIDITY

In proper usage, one is interested in making proper interpretations of a test score, so test makers are responsible for accumulating evidence that support the intended interpretations and uses of the scores (Kane, 2006). In the case of TAKS, EOC, and SDAA II, test results are used to make inferences about students' knowledge and understanding of the TEKS. RPTE assessment results provide both a measure of progress in English language acquisition and information about students' knowledge and understanding of the reading TEKS. This chapter provides evidence about the processes used to develop the tests and the analytic studies conducted to better understand interpretations that may be assigned to individual test scores. The search for validity evidence is a never-ending process, and future technical reports undoubtedly will include additional information in this regard.

Evidence Based on Test Content

Standards-referenced assessments, such as the TAKS, EOC, SDAA II, and TELPAS, are based on an extensive definition of the content they assess. Test validity is, therefore, content based and tied directly to the statewide curriculum. To ensure the highest level of content validity, the process of aligning TAKS, EOC, SDAA II, and TELPAS to the curriculum was carefully implemented and included review by numerous committees of Texas educators.

When TAKS was designed as the standards-referenced assessment for the TEKS, advisory committees consisting of educators from districts across the state were formed for each subject area at each grade level. Teachers, test development specialists, and TEA staff members worked together in these committees to identify TEKS student expectations that were important to assess and to develop test objectives, item development guidelines, and test-item types. In addition, starting in 2001–2002, committees convened to review and edit TAKS items for content and bias and to review data from field testing. A similar process was conducted for EOC, SDAA II, and RPTE when they were developed.

The development of the TAKS Equivalency Standard as part of the SDAA II-to-TAKS linking study provides additional validity evidence for SDAA II. The SDAA II testing program was designed to align with the TAKS and to cover the same range of the TEKS curriculum as the TAKS. The TAKS Equivalency Standard is the score on each SDAA II test given on grade level at which a student demonstrates the same level of performance as a student who met the passing standard on the corresponding TAKS grade-level test. Further information about the TAKS Equivalency Standard can be found in Chapter 14: Standards of this digest and in Appendix 22 of the *2005–2006 Technical Digest* which is located at <http://www.tea.state.tx.us/student.assessment/resources/techdig06/appendices.html>. The fact that the TAKS and SDAA II tests could be successfully linked provides evidence of the content validity of the SDAA II.

Relation to the Statewide Curriculum

The item writers and reviewers for each stage of development verify the alignment of test items with the objectives to ensure that the items measure appropriate content. The sequential stages of item development and item review provide many opportunities for Texas educators to offer suggestions for improving or eliminating items and to offer insights into the interpretation of the statewide curriculum. The nature and specificity of these various review procedures provide additional strong evidence for the content validity of the TAKS, EOC, SDAA II, and RPTE assessments.

Educator Input

Texas educators provide valued input on the content and the connection between the items and the statewide curriculum. Many current and former Texas educators and some educators from other states also work as independent contractors to write items specifically to measure the objectives. These multiple sources of expertise provide for a system of checks and balances for item development and review that reduces single-source bias that might be introduced if items were written by a single author. In other words, because test items are written by many different people with diverse backgrounds, it is less likely that items will suffer from a bias that might occur if items were written by a single author. These multiple sources of expertise provide for a system of direct input from educators offers additional evidence regarding the validity of constructed TAKS, EOC, SDAA II, and RPTE tests.

Test Developer Input

The staff at TEA, as well as professional test developers from Educational Testing Service, Pearson Educational Measurement, and Questar, Inc., provide a wealth of test-building experience, including content expertise. Each internal review of an item by these experts increases the probability of the item being an accurate measure of the intended objective. Hence, these reviews are offered as additional evidence for the content validity of the TAKS, EOC, SDAA II, and RPTE tests.

Test Expert Input

TEA, in conjunction with Pearson Educational Measurement, receives ongoing input from a panel of national testing experts regarding all plans for collecting validity evidence for the Texas assessments. In the case of the state's ELL student assessment procedures, language acquisition and psychometric experts are involved in all phases of assessment development and refinement. Several times a year, Texas convenes an ELL assessment focus group to provide input as a team of language acquisition experts. These experts compose the ELL Focus Group which provides input on technical and practical issues related to the development and refinement of TELPAS and LAT procedures required under the NCLB Act. Issues considered include TOP audit procedures, training plans, and evidence to support TOP's reliability and validity. The ELL assessment focus group is composed predominantly of bilingual/ESL educators and coordinators, district testing coordinators, and campus administrators. Texas has also elicited psychometric input from its Texas Technical Advisory Committee (TTAC) on ELL assessment issues. The TTAC has provided input on several plans for gathering reliability

and validity evidence for TELPAS and has helped shape the plans for Spanish TAKS standard setting and TELPAS audits.

Evidence Based on Relations to Other Variables

Another way to provide validity evidence is by analyzing the relationship between test performance and performance on some other measure. This other measure can be evaluated concurrently or in the future and is then correlated with the test score. In this way, the test score is compared with a criterion that is thought to be a reasonable estimate of the same construct the original test purports to measure. As part of the TAKS Higher Education Readiness Component, a concurrent validity study was conducted in 2004–2005 to correlate performance on exit level TAKS with performance on national testing programs.

TAKS

Criterion-related evidence of validity for TAKS was provided in a study conducted by TEA and PEM to fulfill the Senate Bill 103 requirement that TEA implement a college readiness component as part of the TAKS. The research, called the Higher Education Readiness Component study, included two parts: a contrasting groups study and a performance data correlation study. The contrasting groups study examined the performance of high school juniors on the first administration of the TAKS exit level mathematics and English language arts tests in 2003 as compared to performance on the same TAKS assessments by a sample of second semester college freshmen who had demonstrated college readiness through successful completion of their first semester courses.

The performance data correlation study examined student performance on TAKS in relation to performance on three college readiness measures used statewide for making college readiness and placement decisions: the Texas Academic Skills Program (TASP), the American College Test (ACT), and the Scholastic Assessment Test I (SAT I). The TAKS to TASP, TAKS to ACT, and TAKS to SAT I comparisons incorporated data collected from Texas public high school juniors who took the exit level TAKS and one or more of these other assessments in spring 2003. ACT and SAT I data were also collected for high school juniors who took the TAKS in spring 2004.

Results of the study indicated that the TAKS scale scores at the Met Standard performance level predicted ACT scale scores of approximately 20 for mathematics. Based on a national study of high school graduates from 2002–2004, 50% of students scored at or above this ACT score. The TAKS scale scores at the Met Standard performance level predicted ACT scale scores of approximately 18 for English. Of the high school students in the ACT data, 67% scored at least this high on the ACT English test.

Results of the study also indicated that the TAKS scale scores at the Commended Performance level predicted ACT scale scores of approximately 27 for mathematics. Based on a national study of high school graduates from 2002–2004, 15% of students scored at or above this ACT score. The TAKS scale scores at the Commended Performance level predicted ACT scale scores of approximately 24 for English. Of the high school students in the national study, 29% scored at least this high on the ACT English test.

Results of the study indicated that the TAKS scale scores at the Met Standard performance level predicted an SAT I scale score of approximately 470 for mathematics. Based on a national study of high school graduates, 50% of students scored at or above this SAT I score. The TAKS scale scores at the Met Standard performance level predicted an SAT I scale score of approximately 460 for English. Based on a national study of high school graduates, 50% of students scored at or above this SAT I score.

Results of the study indicated that the TAKS scale scores at the Commended Performance level predicted an SAT I scale score of approximately 620 for mathematics. Based on a national study of high school graduates, 25% of students scored at or above this SAT I score. The TAKS scale scores at the Commended Performance level predicted an SAT I scale score of approximately 580 for English. Based on a national study of high school graduates, 25% of students scored at or above this SAT I score. For further information about the study, see the "Higher Education Readiness Study" in the 2007 Texas Education Agency Technical Report Series which can be found at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.

Another source of criterion-related validity evidence for the TAKS tests is the Grade Correlation Study. This study compared the pass/fail rates of Texas students on the TAKS tests with their passing credit/not passing credit rates in their past related courses. Results indicated that a high percentage of students who pass the TAKS tests also pass their related courses. Small percentages of students passed the TAKS tests but did not pass their related courses, passed their related courses but did not pass the TAKS tests, or failed to pass the TAKS test or their related courses. See the "Grade Correlation Study Report" in the 2007 Texas Education Agency Technical Report Series which can be found at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html> for more details on study.

The transadaption process is an important method used to develop Spanish TAKS reading, mathematics and science. In this process, items for Spanish TAKS were translated from English TAKS and adapted as necessary to ensure linguistic and cultural appropriateness.

A study conducted by Pearson Educational Measurement (Davies, O'Malley & Wu, 2007) on the measurement equivalence of transadapted reading and mathematics tests provided an evidence of validity of TAKS tests. The study investigated the fit of a confirmatory factor model to student data from 2006 for grades 3 and 5 TAKS reading and grades 4 and 6 TAKS mathematics tests across English anchor forms and Spanish transadaptions.

The results of the study indicated good model fit within and across English and Spanish forms for both subjects and grades. The study suggested that the Spanish TAKS and English TAKS items function similarly and that the combination of transadapted and independently developed items on the Spanish TAKS tests is appropriate for assessing both reading and mathematics. The results of the study support that both English and Spanish version of TAKS tests measure the same construct.

Relationship Between RPTE and TAKS Reading/English Language Arts

RPTE is designed to provide both a measure of progress in English language acquisition in reading and information about a LEP student's knowledge and understanding of grade-appropriate TEKS in the content area of reading. One way to evaluate whether the skills assessed on RPTE align with grade-level TEKS expectations in reading is to relate RPTE performance by proficiency level to mean scale score performance on TAKS reading/English language arts tests. The results of this analysis are located in Chapter 4: Texas English Language Proficiency Assessment System (TELPAS).

In the 2004–2005 and 2005–2006 school years, other analyses between RPTE and TAKS were conducted. These analyses were conducted to provide additional evidence concerning the suitability of RPTE as an alternative reading assessment for LEP-exempt recent immigrants in Adequate Yearly Progress (AYP) performance measures.

One set of analyses was conducted to examine how similar the percentages of LEP-exempt students who meet the RPTE incremental progress standard are to the percentages of nonexempt LEP students who pass TAKS. The results of these analyses indicated that the percentages of students who met the RPTE incremental progress standard and TAKS standard were very similar. Specifically, approximately 66% of LEP-exempt recent immigrants met the RPTE incremental progress standard in spring 2004 while approximately 65% of nonexempt LEP students met the TAKS standard. The performance results from the 2005–2006 school year were even more similar. In 2005–2006 the percentage of LEP-exempt recent immigrants who met the RPTE incremental progress standard and the percentage of nonexempt LEP students who met the TAKS standard was the same, 68%. The results of these analyses suggest that the AYP performance hurdle is similar whether the RPTE progress standard or TAKS standard is used. Additional RPTE-TAKS analyses and data are included in the "Relationship between Performance Levels on the RPTE and the TAKS Technical Report" in the 2007 Texas Education Agency Technical Report Series which can be found at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.

A study was also conducted in 2005–2006 which established links between RPTE scores and the TAKS performance categories of Met Standard (passing level) and Commended Performance (highest performance level). TAKS Met Standard concordance scores on RPTE were located near the cut point between the RPTE Advanced and Advanced High proficiency levels. TAKS Commended Performance levels were located nearer the top of the RPTE Advanced High score range. This study, which showed that the range of reading proficiency assessed on RPTE is broad enough to link to both the passing category on TAKS and the commended category, can help school districts gain a better understanding of the way in which high performance on RPTE relates to high performance on TAKS. It also provides validity evidence of the alignment between the reading skills assessed on RPTE and TAKS. The results of this study are consistent with second language acquisition theory in that one would expect a strong relationship between high scores on an English language reading proficiency test aligned to grade-level

reading skills and passing rates on a grade-level criterion-referenced reading achievement test. More information about this linking study, see the "Relationship between Performance Levels on the RPTE and the TAKS Technical Report" in the 2007 Texas Education Agency Technical Report Series which can be found at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.

Gathering Validity Evidence for TAKS–Alt

As with TAKS and SDAA II, TAKS–Alt test results are used to make inferences about students' knowledge and understanding of the TEKS. Unlike other statewide assessments in Texas, TAKS–Alt is not a traditional paper-and-pencil or multiple-choice test. Instead, the assessment involves teachers observing students as they complete instructional activities that link to the grade-level TEKS curriculum. As mentioned in the introductory paragraph of this chapter, validity for this assessment is a process of collecting evidence to support inferences made from the scoring results, but different approaches than those used for TAKS have been used for TAKS–Alt.

Evidence Based on Test Content

Content validity evidence has been collected at all stages of the test development process. Evidence based on test content is information that shows the relationship between content and the construct measured by the test. TAKS–Alt was developed to align with the content defined by TEKS. An explicit mapping of the alignment of the TAKS–Alt with the TEKS can be seen in the following documents developed for TAKS–Alt: the TEKS Vertical Alignment documents, the TEKS Curriculum Framework for the Alternate Assessment documents, and the Example Instructional Activities documents. These documents can be found online at <http://www.tea.state.tx.us/student.assessment/resources/taksalt/>.

Content experts and special education experts have been involved in the development and refinement of the assessment since its inception. Focus groups consisting of teachers and other experts have provided reviews and feedback at multiple points throughout the development process. Educator advisory meetings were convened to provide feedback on the TAKS–Alt participation guidelines, the prototype assessment, and the scoring rubric. Educator review meetings provided teacher feedback on the alignment of the general education curriculum and access activities to the grade level TEKS curriculum and TAKS objectives. Volunteer teachers used the TAKS–Alt scoring rubric and provided feedback. The steering committee convened three times to provide feedback on all aspects of the assessment.

In addition, all internal meetings, steering committee meetings, educator advisory meetings, and educator review meetings are documented. All versions of TAKS–Alt related materials are being maintained, and a technical report is being written to describe the progression of the TAKS–Alt development. The technical report will include information collected during the activities described below.

TAKS–Alt Audit

Additional content validity evidence for the TAKS–Alt was obtained through a series of audits of student responses. Auditors reviewed 10% of the student folders consisting of responses to the four essence statements assessed during the field-test administration. As part of this review, the teacher-developed instructional activities were analyzed to evaluate how well the activities matched the assessed objective and how well the activities matched the assessed essence statement. Since the essence statements were developed to be the core of the knowledge and skill statement under each objective, it was expected that if there was a strong link between the activity and the essence statement, that link would also be present between the activity and the objective. Auditors also reviewed the primary documentation provided by teachers to support the scoring of the instructional activity for each of the four essence statements. Auditors were shown the teacher ratings and evaluated how well the documentation matched the teacher ratings (or student score) for the instructional activity.

Auditors were asked to indicate their level of agreement (strongly agree, agree, disagree, or strongly disagree) with the following three statements for each student folder they viewed:

- 1) The instructional activity is linked to the essence statement being assessed.
- 2) The instructional activity is linked to the objective statement being assessed.
- 3) The documentation supports the student's score.

A total of 1879 student folders were viewed by 124 auditors. In terms of the teacher developed instructional activity being linked to the essence statement, 95% of auditors agreed or strongly agreed with the statement. As expected, 95% of auditors also agreed or strongly agreed with the statement that the instructional activity linked to the objective being assessed. These results provide good validity evidence that the activities teachers are developing to assess students are linked to the content that the teacher planned to assess.

Responses regarding how well the provided documentation supported the student's score were slightly less strong with 75% of auditors agreeing or strongly agreeing with the statement. Further training on student documentation and its link to the student's score may result in more auditors agreeing that the documentation supports the student's score. Training with this focus is currently being developed and will be available for the first operational administration of TAKS–Alt in 2007–2008.

Plans are to repeat the TAKS–Alt audit annually. For the first few of years that the assessment is operational, the audit will follow the above outline. In future years, adjustments may be made to the audit to collect further validity evidence. For example, once teachers have become familiar with providing supporting documentation for the assessment, audit panelists will be asked to rate the student (using the TAKS–Alt rubric) based on the activity description and supporting documentation instead of evaluating how well the documentation matches the teacher ratings. For further information about the TAKS–Alt Audit, see the "TAKS–Alt" report in the 2007 Texas Education Agency Technical Report Series which can be found at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.

Evidence Based on Response Processes

Validity evidence based on response processes involves analysis of the types of responses given by a test taker to examine the fit between those responses and the construct measured by the test. The individualized nature of the TAKS–Alt allows teachers to tailor the assessment to the needs of the student so that the responses the student makes are specific to the instructional activity being conducted.

During the TAKS–Alt pilot study, a modified cognitive lab methodology was used to ascertain how teachers were using the assessment. The modified model included observation and cognitive interviewing. This study used trained observers during instruction-based assessment activities followed by cognitive interviewing with probing questions. The interviews were designed to ensure particular topics were covered but were conversational in tone and structure to encourage free dialogue.

Observations during the cognitive labs provided documentation showing that the behaviors being measured are relevant and important for the special education population. The teachers were consistently knowledgeable about conducting instructional activities, and the activities they designed were age appropriate and linked to the TEKS. Most teachers clearly saw the link between the instructional activity and the associated essence statement.

Interviews conducted with the cognitive lab teachers immediately after completing the assessment indicated how teachers arrived at their ratings using student responses. For further information about the TAKS–Alt cognitive labs see the “TAKS–Alt” report in the 2007 Texas Education Agency Technical Report Series which can be found at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.

Evidence Based on Consequences of Testing

Validity evidence that shows the TAKS–Alt is having a positive impact on student learning and instruction has been collected through teacher surveys after the pilot and field test. Teachers generally agreed that students will be adequately prepared for the 2007–2008 administration. The majority of surveyed teachers were somewhat or very comfortable conducting instructional activities. Additionally, teachers were confident about the rating their student earned. Those teachers that participated in the pilot reported administering the pilot test to their students increased their own understanding of how to administer the field test.

Teachers who participated in both the pilot and the field test were surveyed about changes in student performance from pilot to field test. Very few teachers responded that student performance improved or declined significantly. But when student performance did improve over that time period, the most frequent reason for a marked student improvement was that the student had received more instruction. Other common reasons included that the student had learned the concept and that the presentation of the instructional activity improved. For further information on the TAKS–Alt survey results see the “TAKS–Alt” report in the 2007 Texas Education Agency Technical Report Series which can be found at <http://www.tea.state.tx.us/student.assessment/resources/techdig07/index.html>.