Chapter 10: Scaling

Rationale

The basic score on any test is the raw score, which is the number of items correct, but the raw score alone does not present a wide-ranging picture of test performance, because it can be interpreted only in terms of a particular set of items. When new test forms are administered in subsequent administrations, other types of derived scores should be used to compensate for any differences in the difficulty of the items and to allow direct comparisons of student performance between administrations. The TAAS testing program provides two derived scores that may be useful in describing different aspects of student performance: scale scores and Texas Learning Index (TLI) scores.

Derived scores are computed using the one-parameter Item Response Theory (IRT) or Rasch model. The advantage of using IRT models in scaling is that all the items measuring performance in a particular content area can be placed on the same scale of difficulty. The further value of the one-parameter (Rasch) model over more-complex IRT models is that the Rasch model assumes that for each raw score point, there is only one ability estimate. This relationship allows the Rasch difficulty values for the individual items to be used in computing a Rasch difficulty level for any raw score point on any test constructed from these items.

Scale Scores

For the TAAS writing test at Grades 4, 8, and exit level, the Grade 8 science and social studies tests, the Spanish versions of TAAS (Grades 3–6 reading and mathematics and Grade 4 writing), and the end-of-course tests, scale scores have been developed to ensure that the passing standard is maintained at the same level of difficulty across administrations. However, because the SBOE established the passing standards independently at each grade, direct comparisons of performance across grades should not be made.

The original test form on which the 70%-correct standard was established was calibrated using the Rasch model. This calibration produced a relationship between the raw score and the Rasch achievement score, which was then transformed so that 1500 represented the passing standard. This transformation is reproduced in the formula

$$\text{Scale Score} = \left(\frac{\text{R} - \text{R at Standard}}{\text{Standard Deviation of R}} \right) * 200 + 1500,$$

where *R* is the Rasch achievement level estimate, *R* at Standard is the Rasch achievement level associated with the 70% raw score, Standard Deviation of *R* is the standard deviation of the Rasch achievement level, and 200 and 1500 are the spread and centering constants, respectively. This transformation established the original scale score system, and statistical equating is required to maintain the same level of difficulty for newly developed forms.

The resulting TAAS scale has a range of approximately 400 to 2400, with 1500 corresponding to 70% of the items correct on the first administration of the test when the passing standards were set.

Texas Learning Index

The Texas Learning Index (TLI) was developed to better meet the needs of school districts and students for longitudinal comparability. A metric with two essential characteristics was sought. First, such a metric should provide an index of student achievement toward the goal of passing the exit level TAAS test, which students must pass in order to be eligible to receive a Texas public high school diploma. Second, the metric should permit comparisons between administrations and between grades for use in the accountability system.

The TLI provides a means for schools to be able to demonstrate improvements in their instructional programs even in cases where the passing standard has not yet been met or the passing standard has been exceeded. Likewise, with a derived score such as the TLI, individual students are able to demonstrate improvement regardless of their current achievement relative to the passing standard.

T-SCORE TYPE TRANSFORMATION

The requirements listed above led to the consideration of a vertical scale score system for the TAAS examinations that would place results for Grades 3–8 and the exit level test all on the same scale. TEA convened a panel of measurement experts from across the nation to advise the agency regarding such a scaling. This panel was composed of educators, test publishers, and educational consultants. The committee expressed two main concerns regarding a vertical scaling system. First, placing both Grade 3 and exit level students on the same scale may lead to misinterpretations because of the large difference in the content of the test items at these grades. Second, a vertical scale implies a linear and well-defined curriculum from Grades 3 through exit level when such a well-ordered curriculum may not be in place. The committee concluded that a vertical scale would not meet the needs of TEA and offered an alternative proposal of using a transformed "T-score" type of scale.

A transformed T-score is expressed in terms of standard deviation units away from the mean. For example, if a student earns a raw score of 50 on a test with a mean of 40 and a standard deviation of 10, this student's score is one standard deviation unit above the mean. Traditionally, such a score is referred to as a standard score, or z-score, and can be reproduced with the following formula:

$$\textbf{z-score} = \frac{\textbf{(Observed Score – Mean)}}{\textbf{Standard Deviation}}$$

The student in the example above has an observed score of 50 and a z-score of 1.00. Because standard scores have decimals and typically range from -3.00 to 3.00, an additional transformation is usually made in order to simplify the reporting. A common transformation in which the scale in this example is "re-anchored" to have a mean of 50 and a standard deviation of 10 is often referred to as a T-score transformation. The following formula provides the T-score transformation:

$$\textbf{T-score} = (\textbf{z-score} * 10) + 50$$

Such a transformation simply renames the more-difficult-to-use z-score. After the transformation, the student in the example would have the following scores: a raw score of 50, a z-score of 1.00, and a T-score of 60.

TLI ESTABLISHED AT EXIT LEVEL

The TLI is very much like the T-score previously described. Unlike the T-score, however, the TLI is anchored at the exit level passing standard rather than at the mean of the distribution. To distinguish between the scale score system and the TLI, TEA chose a two-digit metric for the TLI so that it is anchored at the exit level passing standard with a value of 70 and a standard deviation of 15. Because a TLI of 70 represents the passing standard, there is no difference in interpretation between a student who scores 1500 under the scale score system and a student who scores 70 on the TLI scale.

The TLI was first established in spring 1994. The following example illustrates how the TLI was constructed. Suppose the passing standard for the exit level reading test, a scale score of 1500, was equivalent to a z-score of 0.30. That is, the passing standard was 0.30 standard deviation unit above the mean for the group of students taking the exit level test. The TLI transformation therefore established a TLI of 70 at 0.30 standard deviation unit above the mean of the scores for the spring 1994 population taking the exit level tests. Continuing the example using the student scoring 50 items correct, this student would have a TLI of 80.5, or 81:

$$\textbf{TLI} = \left[\left(\left(\frac{\textbf{(Observed Score} - \textbf{Mean)}}{\textbf{Standard Deviation}}\right) - \textbf{z-Score at Passing Standard}\right) * \textbf{15}\right] + \textbf{70}$$

$$\textbf{TLI} = \left[\left(\left(\frac{\textbf{(50} - \textbf{40)}}{\textbf{10}}\right) - \textbf{0.30}\right) * \textbf{15}\right] + \textbf{70} = \textbf{80.5} = \textbf{81}$$

TLI ESTABLISHED AT THE OTHER GRADES

The TLI was also established in spring 1994 at Grades 3–8 to indicate the performance needed in order to be likely to succeed on the exit level test and to allow comparisons of learning progress between grades. Continuing with the example cited in the previous section, it was determined that if the passing standard was 0.30 standard deviation units above the mean for the exit level test, students would need to be 0.30 standard deviation units above the mean of the test at each of the lower grades in order to be in line to reach the passing standard at the exit level. In other words, the same level of relative performance needed to reach a TLI of 70 on the exit level test was also needed to reach a TLI of 70 at each of the other grades.

Note that the TLI is an index of a student's level of performance relative to the mean performance of the other students in the same grade. For example, if a student with a TLI of 70 on the exit level test scored as well as or higher than about 44% of all exit level test takers, a student with a TLI of 70 at Grade 8 also scored as well as or better than about 44% of all Grade 8 test takers. As can be seen in the table on the following page, the TLI associated with the passing standard yielded about the same percentile rank (PR) at all grades within a subject area. The PR was not exactly the same at each grade, because the student scores were discrete in such a way that not all z-scores obtainable at one grade were obtainable at another.

	Mathematics		Reading	
	TLI	PR	TLI	PR
Exit level	70	44	70	25
Grade 8	70	44	70	25
Grade 7	70	43	70	26
Grade 6	70	43	70	28
Grade 5	70	43	70	27
Grade 4	70	43	70	27
Grade 3	70	44	70	24

To construct the TLI, the difficulties of the Grade 3–8 examinations were aligned with the relative difficulty of the exit level exam. The particular subject area test was designed and constructed at each grade so that answering approximately 70% of the items correctly would yield a TLI score of 70. As can be seen in the following table, for the spring 2002 administration, the percent of items needed to meet minimum expectations ranged from 48% to 55% in mathematics and from 60% to 67% in reading. TAAS forms for Grades 3–8 continue to be constructed to be comparable in difficulty with the forms used in the spring of 1994, when the TLI was established.

	Raw Score % Correct	
Grade/Level	**Mathematics**	**Reading**
3	55	67
4	52	63
5	52	65
6	50	65
7	50	64
8	48	60
Exit level	48	60

TLI SUMMARY DATA

Summary statewide performance data for all students not in special education are listed in Appendix 12 of this digest. Appendix 12 provides frequency distributions, central moments, and percentages of students scoring at each TLI score point, as well as the test reliability and the overall standard error of measurement.

TLI SCORE INTERPRETATIONS

The TLI is not a vertical scale, and interpretation of individual TLI differences between grades must be made carefully. Comparisons in growth within a subject area across grades are valid where growth refers to learning progress "typical" for all students between the grades. For example, suppose a student attained a Grade 3 mathematics TLI score of 71. Suppose further that in Grade 4 the student's TLI was also 71. This does not

mean the student did not show growth between Grades 3 and 4. The student's performance is in about the same position relative to fourth graders as the student's performance was relative to third graders. This student "held his or her own" and showed typical learning progress or growth from Grade 3 to Grade 4 by not changing his or her relative rank-ordered position inside each grade distribution.

The TLI defines typical progress as maintaining the same position relative to one's peers (i.e., the relative rank ordering or percentile rank within the grade) from grade to grade. Some differences in rank orderings between the grades will occur because of slight differences in test difficulty, differences in test populations, and measurement error. Hence, a TLI of 70 at Grade 8 does not ensure a TLI of 70 on the exit level test, and continued instruction to realize additional progress must be maintained.

Assumptions

An assumption of the TLI is that the reference distributions on which the TLI scale was constructed should be used for all future TAAS scoring. Because the TLI is a distributional-based metric relying on a z-score transformation, it is normative in nature. Because of this norm-referenced component, recalculating the TLI each year would make year-to-year comparisons impossible. For this reason, *all TAAS administrations will describe student performance in terms of the population tested in the spring of 1994.*

TLI Gains and Losses

Statewide average TLI gains and losses from the TAAS administrations in spring 2001 and spring 2002 in reading and mathematics for Grades 3–8 follow this text. Gains and losses are shown for all students not in special education, including students who did or did not meet minimum expectations and students who scored at specific points in the spring 2002 distribution.

Based on the design and structure of the TLI, there would be little or no gain expected between grades if students were making typical progress across the years—that is, if student performance was neither improving beyond nor falling behind expectations. For instance, a student scored a TLI of 75 in the spring of 2001 and continued to learn at the same rate as his or her peers should have scored at or near a TLI of 75 in the spring of 2002 because the TLI was constructed so that making typical progress from year to year results in the same TLI score. If substantial changes in gains or losses should occur, there are several pieces of information that may be examined to help determine whether these gains or losses were actual changes in achievement.

The overall gains presented in the following tables indicate that between 2001 and 2002, students demonstrated more than typical learning progress, with "typical" progress defined as little or no gain in TLI from grade to grade. Furthermore, the tables show that students who failed TAAS in 2001 made significant gains in their TLI scores, while students who passed TAAS in 2001 exhibited little change in TLI.

The tables also provide the TLI gains or losses exhibited in the spring 2002 TAAS administration for students who scored at specific points in the spring 2001 distribution. For example, on the TAAS reading test at Grade 3, students who scored at the 25th percentile of the spring 2001 TLI with a TLI of 3-79 made an average TLI gain of 4.9 on the spring 2002 administration. By examining these gains or losses at specific points in the distribution, a better understanding of the nature of these gains and losses may be achieved.

Average TLI Gain/Loss at Adjacent Grades
All Students Not in Special Education
Spring 2001–Spring 2002

READING

	Grade 3 to 4		Grade 4 to 5		Grade 5 to 6		Grade 6 to 7		Grade 7 to 8	
	2001 TLI Score	2002 Gain/Loss	2001 TLI Score	2002 Gain/Loss	2001 TLI Score	2002 Gain/Loss	2001 TLI Score	2002 Gain/Loss	2001 TLI Score	2002 Gain/Loss
Overall	–	4.2	–	2.4	–	0.2	–	2.7	–	2.9
Failing TAAS	–	12.2	–	11.6	–	8.6	–	11.2	–	13.2
Passing TAAS	–	3.2	–	1.6	–	–0.6	–	1.5	–	1.9
Minimum	3-2	81.7	4-15	68.4	5-8	76.9	6-18	63.2	7-16	69.3
P10	3-69	9.3	4-73	6.4	5-71	3.5	6-67	9.5	7-73	8.0
P25	3-79	4.9	4-83	2.8	5-84	0.6	6-80	4.7	7-84	4.0
Mean	3-84	3.4	4-87	2.1	5-88	–0.1	6-86	3.2	7-88	3.0
P50 (Median)	3-86	4.0	4-90	1.6	5-92	–0.5	6-89	1.6	7-91	2.0
P75	3-91	2.2	4-96	0.1	5-96	–1.3	6-95	–0.3	7-96	–0.1
P90	3-94	1.8	4-98	–0.6	5-98	–1.9	6-98	–2.2	7-98	–1.3
Maximum	3-94	1.8	4-98	–0.6	5-101	–3.6	6-100	–3.3	7-99	–1.6

NOTE: P denotes percentile (i.e., P10 is the 10th percentile).

Average TLI Gain/Loss at Adjacent Grades
All Students Not in Special Education
Spring 2001–Spring 2002

MATHEMATICS

	Grade 3 to 4		Grade 4 to 5		Grade 5 to 6		Grade 6 to 7		Grade 7 to 8	
	2001 TLI Score	2002 Gain/Loss	2001 TLI Score	2002 Gain/Loss	2001 TLI Score	2002 Gain/Loss	2001 TLI Score	2002 Gain/Loss	2001 TLI Score	2002 Gain/Loss
Overall	–	2.8	–	3.6	–	0.0	–	0.7	–	0.9
Failing TAAS	–	14.4	–	11.6	–	7.6	–	6.5	–	9.3
Passing TAAS	–	0.9	–	3.0	–	–0.4	–	0.3	–	0.2
Minimum	3-1	78.7	4-9	74.1	5-10	70.0	6-14	64.6	7-13	68.2
P10	3-66	11.4	4-73	7.3	5-75	2.9	6-74	1.8	7-72	4.1
P25	3-77	4.9	4-80	4.4	5-82	0.7	6-81	0.3	7-80	1.0
Mean	3-81	1.9	4-83	3.3	5-85	–0.3	6-84	–0.4	7-84	0.4
P50 (Median)	3-84	1.1	4-85	2.8	5-87	–0.3	6-87	0.1	7-86	0.0
P75	3-89	–1.4	4-88	1.6	5-90	–1.0	6-90	0.4	7-90	–0.6
P90	3-91	–2.5	4-89	1.6	5-92	–1.9	6-91	0.3	7-92	–1.2
Maximum	3-93	–3.5	4-91	0.8	5-93	–2.3	6-92	–0.1	7-93	–1.6

More Information

- Conversion Tables

  The table in Appendix 11 provides information for converting raw scores to derived scores for all grades and content areas assessed.

- Frequency Distributions and Descriptive Statistics

  Appendix 12 provides frequency distributions and summary statistics for the TLI scales.