2008 TAKS English Vertical Scaling Study Report

April 4, 2009

Prepared for the Texas Education Agency
by
Pearson
Psychometric and Research Services

Table of Contents

Background

Tracking students' academic progress from year to year and over the course of their schooling has become increasingly important for large-scale assessment programs. The implementation of the No Child Left Behind Act of 2001 (NCLB; Public Law 107-110) requires annual assessment of student progress in the state curriculum standards at designated grade-based intervals. Student scores on such assessments can be represented as numerical values that can be used to compute and describe differences in student performance within a given grade level. This scaling technique is termed *horizontal scaling*. Because horizontal scales are designed to represent within-grade level differences, they cannot be used to express student growth across grades. If the assessment of student performance requires the monitoring of student growth, then the expression of student performance can benefit from the association of student scores with numerical values that reflect across-grade level performances. This scaling technique is called *vertical scaling*.

Since its inception in 2002-2003, the Texas Assessment of Knowledge and Skills (TAKS) has used a horizontal scale. However, under Section 39.036 in S. B. No. 1031, the Texas Educational Agency (TEA) is required to develop a vertical scale for assessing student performance beginning in the 2008-2009 school year in the following areas:

- reading, grades 3 through 8
- mathematics, grades 3 through 8

Data for a vertical scale score system were collected during spring 2008 so that a vertical scale score system can be in place by the 2009 administration.

Data Collection Design

The data collection design for the TAKS vertical scaling study used embedded field-test positions on the assessments during the operational 2008 TAKS administrations. This design required designating either four or eight of the regular field-test forms for the vertical scaling study. Using the embedded field-test positions was desirable because students would have no knowledge of whether an item was an operational or vertical scaling anchor item.

The data collection model was based on a common-item nonequivalent group design with dual grade common-items for both the TAKS reading and mathematics tests. An example of this data collection is depicted in Figure 1 below. The model required using (a) only previously field-tested items as anchor items, (b) both lower grade-level and upper grade-level anchor items; and (c) embedded field-test positions. For example, some of the previously field-tested grade 3 items from the TAKS item bank were used as

anchor items and were placed on both the grade 3 and grade 4 vertical linking forms. Likewise, some of the previously field-tested grade 4 items from the item bank were used as anchor items on grade 3, grade 4, and grade 5 vertical linking forms. The number of vertical linking forms needed for this model is eight for grades 4-7. Because the anchor items are only from one adjacent off-grade level, grades 3 and 8 only have four forms each.

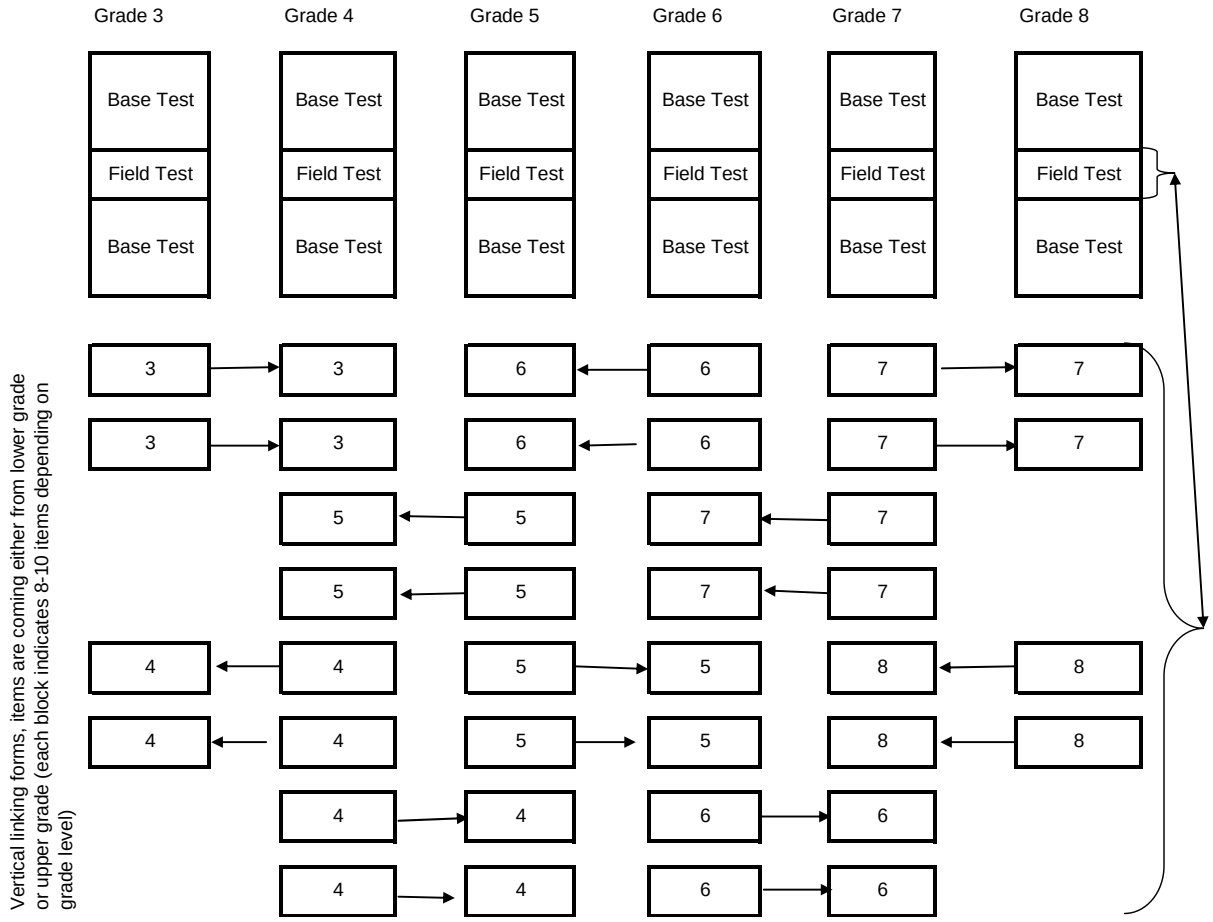


Figure 1. TAKS English Reading and Mathematics Vertical Scale Data Collection Design.

TAKS English Vertical Scale Anchor Forms Descriptive Statistics

Vertical linking items used the same locations as field-test items. Table 1 displays the number of vertical linking items between adjacent grade levels as well as the number of base test items for each grade level. Table 2 displays the sample size for the reading anchor forms by grade level. Between 4,800 and 6,100 students per form were part of the reading vertical scale analyses. Table 3 displays the sample size for the mathematics anchor forms by grade level, with the sample sizes ranging from nearly 5,000 to over 6,000 students per form.

Table 1. Possible number of TAKS vertical linking items by grade levels.

		TAKS Vertical Scale Grades					
	Direction	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Read	Base	36	40	42	42	48	48
	Lower	--	20	20	20	20	20
	Upper	20	20	20	20	20	--
Math	Base	40	42	44	46	48	50
	Lower	--	16	16	20	20	20
	Upper	16	16	20	20	20	--

Table 2. Sample size of reading vertical scale forms

FORM	Gr3	GR4	GR5	GR6	GR7	GR8
Overall	19416	45462	39437	48131	48454	19308
2	4830	5612	4914	6009	6029	4849
3	4893	5656	4924	5998	6041	4815
4	4851	5736	4925	6015	6023	4816
5	4842	5686	4918	6006	6068	4828
6	--	5681	4927	6038	6077	--
7	--	5680	4920	6035	6096	--
8	--	5691	4963	6021	6052	--
9	--	5720	4946	6009	6068	--

Table 3. Sample size of mathematics vertical scale forms

FORM	Gr3	GR4	GR5	GR6	GR7	GR8
Overall	25528	45876	39610	48205	48573	19421
2	6360	5662	4916	6033	6059	4873
3	6386	5724	4930	6018	6071	4847
4	6389	5780	4943	6022	6029	4850
5	6393	5752	4969	6022	6076	4851
6	--	5743	4992	6019	6085	--
7	--	5734	4963	6045	6111	--
8	--	5733	4961	6024	6062	--
9	--	5748	4936	6022	6080	--

Main Steps to Developing the TAKS English Reading and Mathematics Vertical Scales

1. Calibration of Vertical Linking Items

To develop a common scale across grade levels, vertical linking items were calibrated both on-grade level and off-grade level with the live (i.e., operational) items. For example, grade 3 items that appeared on both a grade 3 and a grade 4 form would be calibrated with the live items from the the grade 3 test (on-grade level) and grade 4 test (off-grade level). Based on feedback from the Texas Technical Advisory Committee (TTAC), within-grade calibration of vertical linking items and live items was done with separate (rather than concurrent) calibrations. The vertical linking study was conducted without linking to the TAKS base scale so that an estimate of across grade level changes in difficulty would not be contaminated by horizontal equating effects.

A two-step calibration procedure was used.

1) The first step was to calibrate only on-grade level live items together using all available students and to compute a mean ($\overline{b}_{On-grade\_only}$) of the resulting Rasch item difficulty estimates.
2) In the second step, the researcher calibrated both on-grade level live items and the vertical linking items (which includes both on-grade level and off-grade level items) together and computed an equating constant by computing the mean difference in on-grade level item parameters between the first step calibration ($\overline{b}_{On-grade\_only}$) and the second step calibration($\overline{b}_{On-grade\_withVS}$) as shown in equation (1) below.

$$C = \overline{b}_{On-grade\_only} - \overline{b}_{On-grade\_withVS} \quad (1)$$

This mean difference (C), the mean/mean equating constant (Kolen & Brennan, 2004), was applied to all item difficulties. This calibration method was done form by form for all vertical linking forms.

2. Diagnostic Information and Item Level Decisions

Model Misfit

Because the vertical scale is built using the Rasch model, it is important to detect any vertical linking item that does not fit this model. The Rasch Mean Square infit statistic was used to identity any misfiting items. If the Rasch Mean Square infit was unexpectedly large or small, then the item was eliminated from the vertical linking set at grade level and both adjacent grade levels. There were only a few items that fell into these categories (3 items in reading and 4 in mathematics, see Table 4 for more information), which had either low (<0.80) or high (>1.20) mean square infits.

Table 4. Items eliminated during model misfit analysis

Grade	Reading	Math
03	0	1
04	0	0
05	1	0
06	1	1
07	0	1
08	1	1
Total	3	4

Outlier Analyses

Another indicator of model misfit was outliers. A two-step process was used to identify outliers. First, a regression analysis was run using the on-grade level Rasch item difficulties (RID) and off-grade level RID. Although a visual inspection of regression plots, such as that shown in Figure 2, is a common method of outlier detection, it was decided to first use the externally studentized residual. An externally studentized residual is the residual for the fitted regression, excluding the current case under observation, divided by its error variance (which is also calculated excluding the current case under observation). An item was eliminated if the absolute value of its externally studentized residual was greater than 2.0, which was chosen because the studentized residual is distributed as a t-statistic. Second, the remaining Rasch item difficulties for the vertical linking items from on-grade and off-grade clusters were plotted against each other and then compared to a graph of items before any outlier was removed. The visual inspection process had two goals. The first was to verify that use of the externally studentized residuals was beneficial. Second, the visual inspection was used to eliminate any items that were missed by the externally studentized residual analysis but clearly needed to be removed. In all, a total of four items were eliminated using visual inspection.

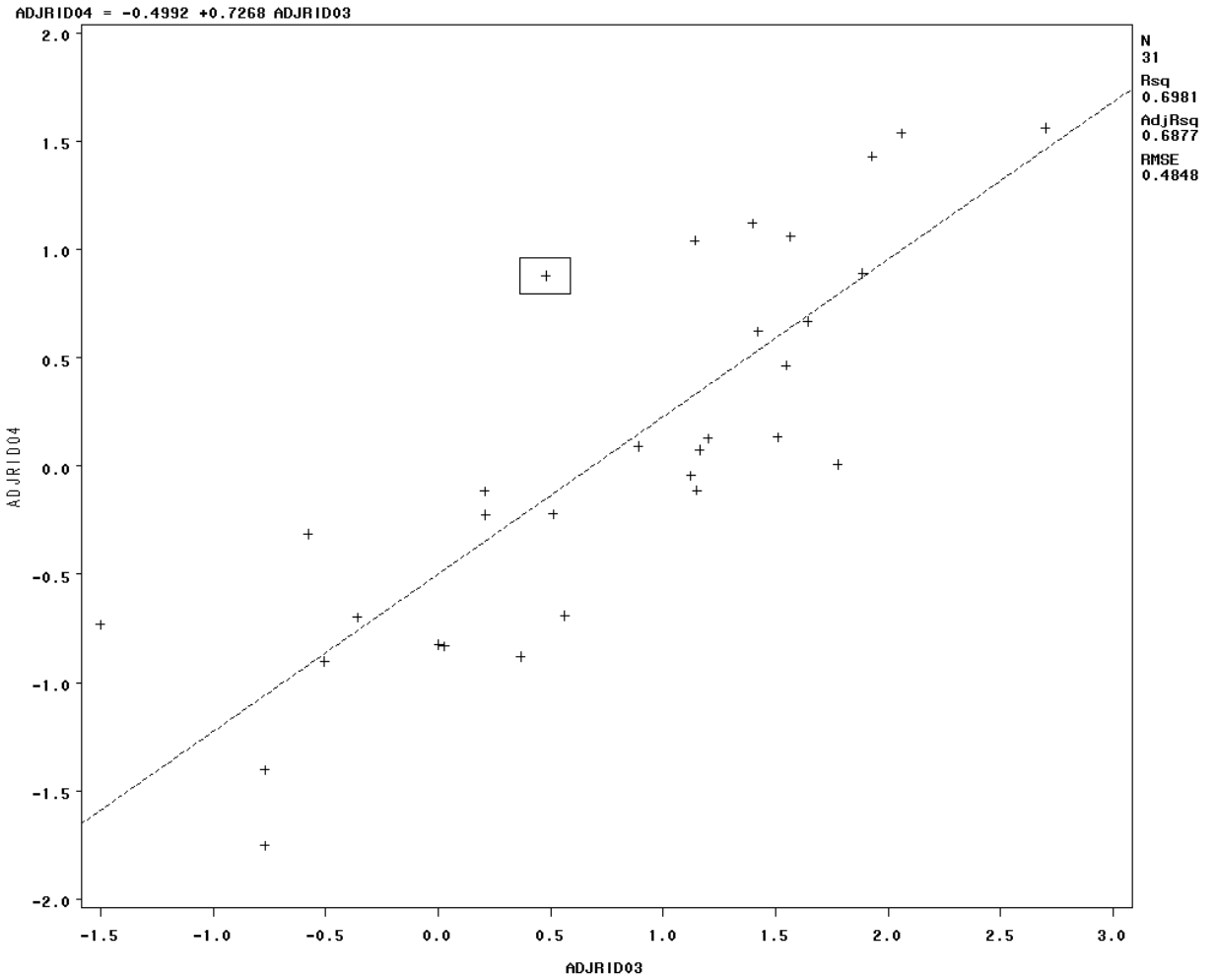


Figure2. Example plot used during visual inspection

Content Representation

It is not only important to inspect any vertical linking item set for items that are not functioning as expected, it is also important to review the removal of any vertical linking items in the context of its impact on the creation of the vertical scale. Table 5 provides an initial count of the vertical linking items by grade set, the number of items eliminated at each step of the diagnostic analyses, and provides the final number of items used to create the vertical scales.

Because the actual number of items eliminated does have an impact on the ability to create the vertical scale, it is also vital that the final vertical linking item set covers the content as was originally intended. Table 6 displays, by objective, the percentage of reading items, for each adjacent grade set, that were in the initial vertical linking item set and the percentage in the final vertical linking item set after the removal of misfiting items and outliers. After reviewing a similar table for mathematics, the final content

representation of the mathematics 3/4 vertical linking item set did not match the initial content representation as well as we would like. Initially, all (n=2) the grade 3 common items were eliminated for the objective titled Concepts and Uses of Measurement. One of the items was eliminated due to a Rasch Mean Square infit larger than 1.20, whereas the other item was eliminated because its externally studentized residual was greater than 2.0. The addition of one of the eliminated items back into the vertical linking set would sufficiently satisfy the content representation for this objective. Furthermore, it was decided that given the use of the Rasch model, the addition of an item with an externally studentized residual greater than 2.0 was better than the addition of an item that did not fit the Rasch model. As a final verification, the correlation between the grade 3 and grade 4 Rasch item difficulties for the mathematics 3/4 vertical linking item set before and after the inclusion of the item was examined. The correlation changed by 0.05. Table 7 displays, by objective, the percentage of mathematics items for each adjacent grade set that were in the initial vertical linking item set and the percentage in the final vertical linking item set.

Table 5. Comparison of vertical linking items sets before and after diagnostic analyses

		Initial			After Infit		After Outlier		Final		
		TOTAL	LOWER	UPPER	LOWER	UPPER	LOWER	UPPER	TOTAL	LOWER	UPPER
READ	03/04	39	20	19	0	0	2	0	37	18	19
	04/05	40	20	20	0	1	1	0	38	19	19
	05/06	40	20	20	0	1	1	0	38	19	19
	06/07	40	20	20	0	0	0	2	38	20	18
	07/08	40	20	20	0	1	2	1	36	18	18
MATH	03/04	32	16	16	1	0	3	2	26	12	14
	04/05	32	16	16	0	0	2	1	29	14	15
	05/06	32	16	16	0	0	0	2	30	16	14
	06/07	40	20	20	1	1	3	1	34	16	18
	07/08	40	20	20	0	1	1	1	37	19	18

Table 6. Percentage of reading vertical linking items by objective

Objective	03/04		04/05		05/06		06/07		07/08	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final	Initial	Final
01	36%	35%	30%	26%	28%	29%	30%	32%	30%	33%
02	18%	19%	13%	13%	23%	24%	10%	11%	18%	17%
03	18%	19%	30%	32%	25%	24%	23%	21%	23%	19%
04	28%	27%	28%	29%	25%	24%	38%	37%	30%	31%

Table 7. Percentage of mathematics vertical linking items by objective

Objective	03/04		04/05		05/06		06/07		07/08	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final	Initial	Final
01	25%	27%	28%	31%	19%	20%	25%	26%	18%	19%
02	13%	15%	16%	10%	16%	17%	20%	15%	23%	24%
03	13%	12%	6%	7%	16%	17%	10%	9%	20%	19%
04	13%	12%	13%	10%	16%	10%	10%	12%	10%	11%
05	13%	12%	13%	14%	16%	13%	15%	18%	10%	8%
06	25%	23%	25%	28%	22%	23%	20%	21%	20%	19%

3. Finding Vertical Linking Constants for Adjacent Grades

The Mean/Mean method (Kolen & Brennan, 2004) was used to define vertical equating constants between the grade levels. Vertical linking constants were computed in three ways:

1) using only lower grade-level vertical linking items
2) using only upper grade-level vertical linking items
3) using a combined set (both lower and upper grade-level vertical linking items).

For example, the *vertical linking constant* between grades 6 and 7 would be the difference between the mean of the vertical linking Rasch item difficulties for grade 6 and grade 7. This process is depicted in Table 8. Tables 9 and 10 provide the vertical linking constants for TAKS reading and mathematics.

Table 8. Calculation of vertical linking constants for TAKS reading and mathematics

Grade	Vertical Linking Constant
0304	$VC_{34} = VL_3 - VL_4$
0405	$VC_{45} = VL_4 - VL_5$
0506	$VC_{56} = VL_5 - VL_6$
0607	$VC_{67} = VL_6 - VL_7$
0708	$VC_{78} = VL_7 - VL_8$

Table 9. TAKS reading vertical linking constants by vertical linking item types

Adjacent Grade Levels	Vertical Linking Constants Using Only Lower Grade Items	Vertical Linking Constants Using Only Upper Grade Items	Vertical Linking Constants Using All Items
0304	1.0531	0.9728	1.0119
0405	0.7786	0.7581	0.7683
0506	0.1630	0.2028	0.1829
0607	0.4282	0.4607	0.4436
0708	0.0320	-0.0424	-0.0052

Table 10. TAKS mathematics vertical linking constants by vertical linking item types

Adjacent Grade Levels	Vertical Linking Constants Using Only Lower Grade Items	Vertical Linking Constants Using Only Upper Grade Items	Vertical Linking Constants Using All Items
0304	0.7850	0.8573	0.8239
0405	0.5424	0.7378	0.6435
0506	0.7480	0.8405	0.7973
0607	0.7036	0.7704	0.7350
0708	0.3572	0.3405	0.3490

Estimating vertical linking constants using three methods provided options to create the vertical scale in different ways. To quantify the difference in results between using the lower and upper grades vertical linking item sets for each subject, a Cohen's *d* was calculated. Because the number of vertical linking items sometimes differed between the lower and upper grades, a weighted, pooled average standard deviation was used. Cohen (1998) provided general guidelines for interpreting *d*. A *d* above an absolute value of 0.80 is considered large. A *d* between an absolute value of 0.50 and 0.79 is considered medium. A *d* between an absolute value of 0.20 and 0.49 is considered small. Table 11 provides the Cohen's *d* between the upper grade vertical linking constants and lower grade vertical linking constants for reading and math. Only one difference, the 4/5 vertical linking item set for mathematics, would be considered to have a practical difference.

Table 11. Cohen's *d* between the upper grade vertical linking constants and lower grade vertical linking constants for reading and math

Adjacent Grade Levels	Reading	Math
03/04	-0.103	0.082
04/05	-0.028	0.212
05/06	0.063	0.099
06/07	0.045	0.076
07/08	-0.095	-0.018

4. Defining the Base Grade Cluster

The base grade refers to the anchoring point of the vertical scale and its determination is a largely arbitrary decision. Because most of the focus will be given to student progress towards passing in grade 8, the base grade for the TAKS English vertical scales was set at grade 8.

5. Computation of the Final Vertical Linking Constant

As indicated earlier, the Mean/Mean equating procedure was used to find the linking constants between adjacent grade levels. After finding the vertical linking constant between adjacent grades, a cumulative linking constant was defined from the base grade to the lower grade levels for any grade level that was not adjacent to the base grade. For example, at grade 6 the *vertical linking constant* (between grades 6 and 7) would be the difference between the mean of the vertical linking RID for grade 6 and grade 7. On the other hand, the *cumulative vertical linking constant* (between grade 6 and the base grade level, grade 8), is the vertical linking constant between grades 7 and 8 *minus* the vertical linking constant between grades 6 and 7 (see Table 12 below). Tables 13 and 14 provide the cumulative vertical linking constants for TAKS reading and mathematics.

Table 12. Calculation of final cumulative vertical linking constants for TAKS reading and mathematics

Grade	Final Vertical Linking Constant
3	$0\text{-}VC_{78}\text{-}VC_{67}\text{-}VC_{56}\text{-}VC_{45}\text{-}VC_{34}$
4	$0\text{-}VC_{78}\text{-}VC_{67}\text{-}VC_{56}\text{-}VC_{45}$
5	$0\text{-}VC_{78}\text{-}VC_{67}\text{-}VC_{56}$
6	$0\text{-}VC_{78}\text{-}VC_{67}$
7	$0\text{-}VC_{78}$
8	0

Table 13. Final cumulative vertical linking constants for TAKS reading

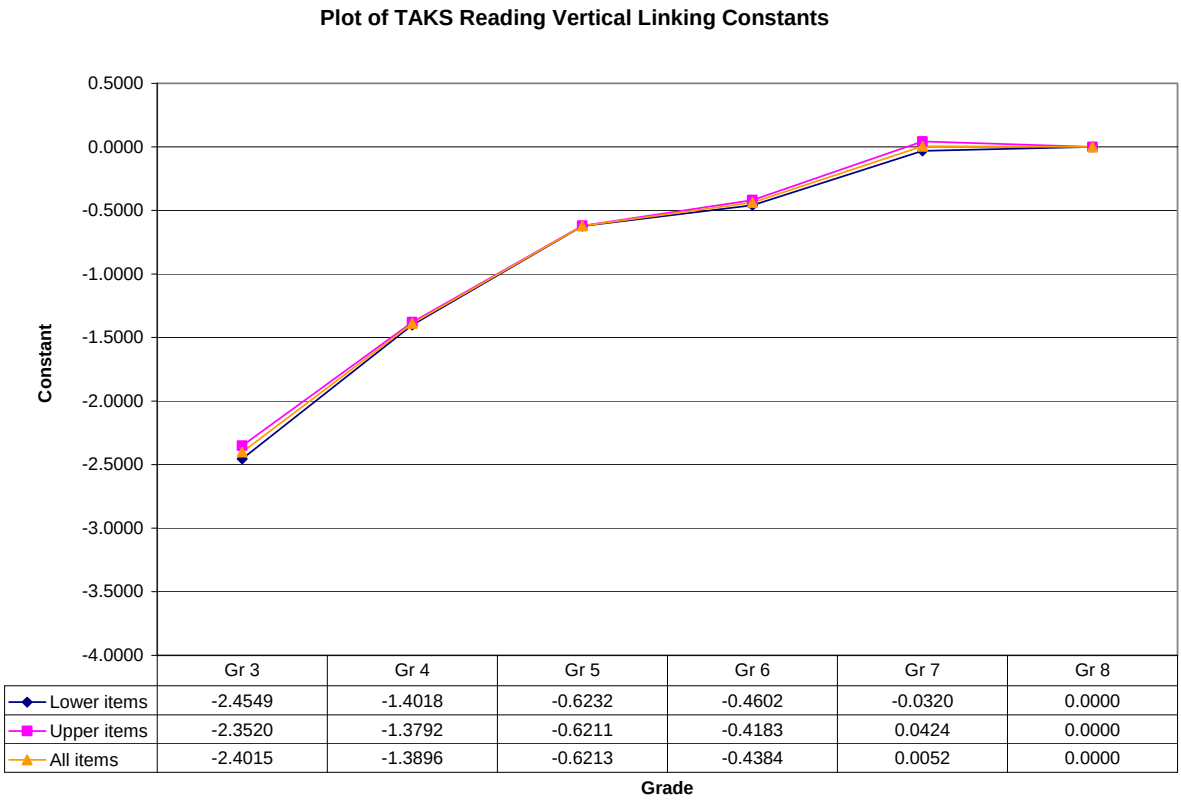
Grade Levels	Final Cumulative Vertical Linking Constants Using Only Lower Grade Items	Final Cumulative Vertical Linking Constants Using Only Upper Grade Items	Final Cumulative Vertical Linking Constants Using All Items
03	-2.4549	-2.3520	-2.4015
04	-1.4018	-1.3792	-1.3896
05	-0.6232	-0.6211	-0.6213
06	-0.4602	-0.4183	-0.4384
07	-0.0320	0.0424	0.0052
08	0.0000	0.0000	0.0000

Table 14. Final cumulative vertical linking constants for TAKS mathematics

Grade Levels	Final Cumulative Vertical Linking Constants Using Only Lower Grade Items	Final Cumulative Vertical Linking Constants Using Only Upper Grade Items	Final Cumulative Vertical Linking Constants Using All Items
03	-3.1362	-3.5464	-3.3488
04	-2.3512	-2.6891	-2.5249
05	-1.8088	-1.9513	-1.8814
06	-1.0608	-1.1108	-1.0841
07	-0.3572	-0.3405	-0.3490
08	0.0000	0.0000	0.0000

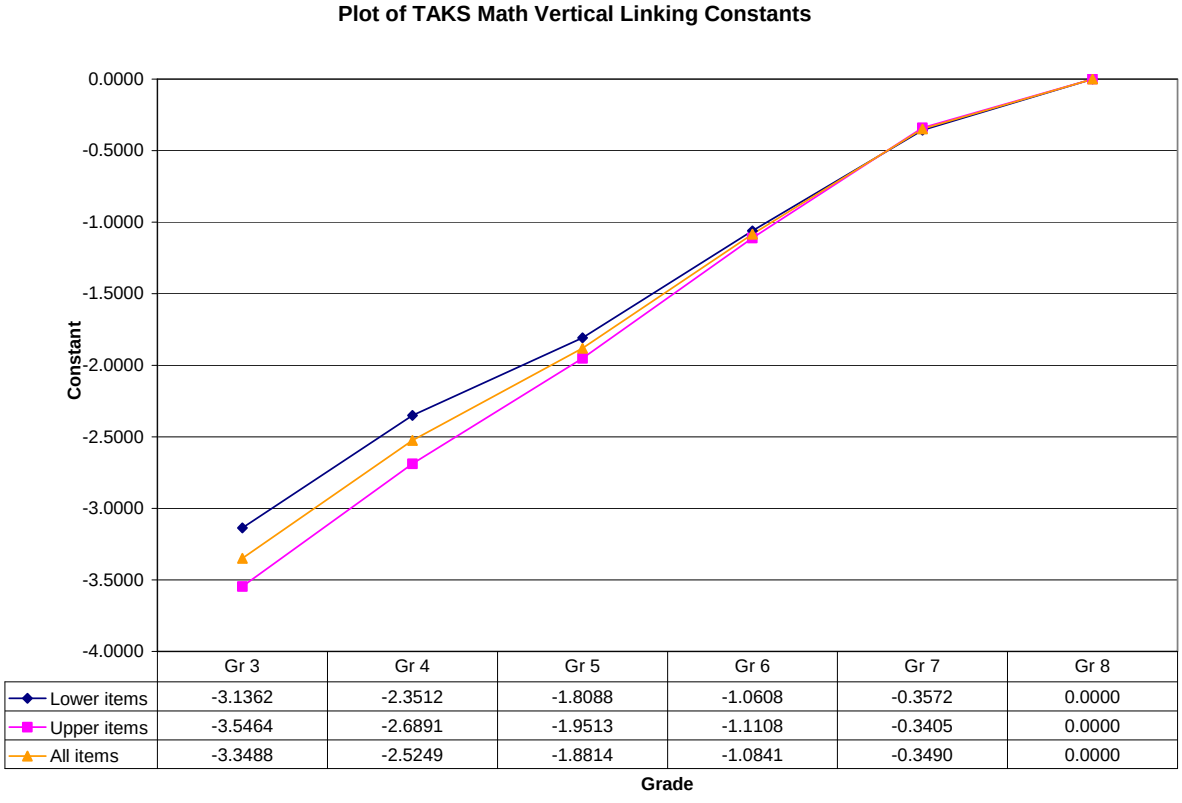
6. Selection of the Final Vertical Scale

The final vertical equating constants were plotted in Figure 3 (reading) and Figure 4 (mathematics). Although all three methods produced similar results, the final vertical equating scale computed from all items will be more stable. As such the “all items” scale is recommended. The final vertical linking constants that result from the selected vertical scale will be used for future TAKS English grade 3-8 reading and mathematics administrations.



	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8
Lower items	-2.4549	-1.4018	-0.6232	-0.4602	-0.0320	0.0000
Upper items	-2.3520	-1.3792	-0.6211	-0.4183	0.0424	0.0000
All items	-2.4015	-1.3896	-0.6213	-0.4384	0.0052	0.0000

Figure 3. TAKS English reading final vertical linking constants



	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8
Lower items	-3.1362	-2.3512	-1.8088	-1.0608	-0.3572	0.0000
Upper items	-3.5464	-2.6891	-1.9513	-1.1108	-0.3405	0.0000
All items	-3.3488	-2.5249	-1.8814	-1.0841	-0.3490	0.0000

Figure 4. TAKS English mathematics final vertical linking constants

7. Customization of Vertical Scale and Smoothing

It might be necessary to place the newly developed vertical scale on another scale with properties that are easier to communicate to stakeholders (e.g., parents, teachers, and school administrators). At the same time, the vertical scale may need to be smoothed to eliminate any roughness in the scale from one grade level to another grade level. Smoothing allows for a more uniform interpretation across grade levels. The Texas Education Agency has decided not to smooth the TAKS English reading and mathematics vertical scales.

8. Evaluation of Vertical Scale

The evaluation of a vertical scale is difficult. There are no firm guidelines by which one can judge all such scales. However, for the TAKS English vertical scales, several criteria have been laid out for evaluating the scale. The development of these criteria was in part based on consultation with Dr. Michael Kolen and from a recent paper by Patz (2007). The following describe the TAKS vertical scale evaluation criteria:

1) Progression in difficulty across grades

Test-level Progression

It seems a reasonable assumption that, when comparing tests designed to assess content at different grade levels, the difficulty of the upper grade test would be higher than the difficulty of the lower grade test. Therefore, an initial check of the TAKS vertical scales was made to verify that this progression exists. Figures 3 and 4 indicate that the vertical scales for both reading and mathematics progress show upward trends, indicating that the average difficulty of the upper grade test is higher than the previous grade. For example, the difficulty of the grade 4 test is higher than the difficulty of the grade 3 test.

Item-level Progression

Another reasonableness check on a vertical scale is the performance of the vertical linking items. At the item level, it seems reasonable that an item should not perform dramatically differently, relative to the other items, across grades. One way to look at across grade differences is to examine the correlation coefficient between RIDs for items in a grade set. As Patz (2007) noted, “high degrees of correlation suggest that the examinees and/or items would be ordered the same way on adjacent test levels, which may be taken as a degree of validation that the vertical scale is appropriate.” (p.18) Table 15 provides the correlation between RIDs for the final adjacent grade vertical linking item sets. Across reading and mathematics, the correlations were high and positive, with the lowest being 0.86 in the 3/4 set for mathematics and the highest being 0.99 for both the 6/7 and 7/8 set for reading.

Table 15. Correlation of item RIDs for the adjacent level final vertical linking item sets

Adjacent Grade Levels	Reading	Math
03/04	0.96	0.86
04/05	0.98	0.95
05/06	0.96	0.92
06/07	0.99	0.97
07/08	0.99	0.97

2) Scale Score means should increase across grade levels in a regular pattern

Just as the vertical linking constants should progress in difficulty, it is reasonable to assume that the application of these constants should affect the population of interest in the same manner. Table 16 displays the means and standard deviations for the vertically scaled theta values for the scores of the students who participated in the vertical scaling study. Theta values were used because scale scores have not been assigned to the assessment yet. As expected, the means of the vertically scaled thetas increased across grade level and the increase formed a regular pattern.

Table 16. Mean and standard deviation of vertically scaled theta values across grade levels

	Reading		Math	
Grade	Mean	SD	Mean	SD
03	-0.072	1.462	-1.348	1.331
04	0.780	1.335	-0.391	1.331
05	1.390	1.214	0.268	1.375
06	1.876	1.256	0.739	1.391
07	2.184	1.244	1.003	1.318
08	2.618	1.242	1.410	1.268

3) Scale Score SDs should not have large differences from grade to grade

As shown in Table 14, the standard deviations of the vertically scaled thetas are quite stable with some small decrease as the grade level increases.

4) The relationship between lower, upper, and combined vertical linking item sets should be regular (i.e., vertical scale line plots from each group do not cross) and the differences between the vertical scales derived from each vertical linking item set should be minimal (plot of final vertical scaling constants)

Estimating vertical equating constants in three different ways provides options for creating the vertical scale. This also provides a good way to cross-validate the vertical scaling methodology. If the different vertical

linking item sets provide the same estimate of growth then the vertical scaling constants should be very similar. If they provide different estimates of growth, then one would expect that differences would be approximately a constant. Figures 3 and 4 can be used to evaluate this criterion. For reading (figure 3), it is easy to see that this criterion is met. The line graphs for the different vertical scaling constants are virtually identical. For mathematics (figure 4), the lines are very similar except for the departure that begins at grade 3 and ends at grade 5. However, the pattern remains constant in those grades.

5) The vertical scale should be reasonably close to the relationship between the grade level TEKS

To evaluate this criterion, separate mathematics and reading meetings were held with content teams from TEA and Pearson. The meetings were broken into three segments: a brief overview of vertical scaling, discussion of the growth expectations given the TEKS and knowledge about the student population, and a review of the actual vertical scale for the subject area.

For reading, the content teams expected to see growth throughout grades 3-8. However, the expectation was that the growth would be greatest between grades 3 and 4 due to the differences in the TEKS for these grades. The content teams also expected that the growth between grades 4 and 5 would also be strong. In the middle grades, the content teams expected some growth but this was harder to predict because the middle school TEKS are the same for grades 6-8. The actual vertical scale is similar to the expectations of the content teams. The one concern for the content teams was the lack of growth between grades 7 and 8. However, the content teams indicated that this is reasonable given the similarities between the grade 7 and 8 reading tests.

In discussions with the mathematics content teams, the expectation was that the vertical scale should show fairly steady growth from grades 3-5. The growth would begin to slow in middle school, especially between grades 7-8. These expectations matched the results of the vertical scale for mathematics.

Given the defined criteria, the 2008 TAKS English vertical study resulted in a reasonable scale that can be used to monitor students' growth as they progress through their various instructional levels in Texas.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ. : Erlbaum.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.

Patz, R. (2007). *Vertical scaling in standards-based educational assessment and accountability systems.* Paper presented at the Council of Chief state school officers.

Appendix A. Vertical Linking Item Selection Guidelines

Vertical Linking Item Selection Guidelines
08/08/2007

Content Expectation of Vertical Linking Items

- Vertical linking items should have already been field tested.
- It is preferable to use items field tested within the most recent three years.
- No griddable items would be used for the vertical linking.
- Choose items that do not require extra materials such as a ruler or calculator when used on an off-grade level test.
- Vertical linking items should be representative of the overlapping content from an adjacent grade.
- All of the vertical linking items from one grade level can be considered to be a grade-level mini test.
- Upper grade level vertical linking items should come from a content area where off-grade students had exposure to the topic, especially in mathematics.
- Passage-based off-grade level vertical linking items on a vertical linking form should come from one passage.
- Selected upper grade level passages should be as close as possible to the off-grade level requirements in characteristics such as word counts etc.
- Vertical linking passages should reflect their grade level passage types.

Psychometric Properties of vertical linking items

- Avoid extremely easy or difficult items such that item difficulty (p-value) range should be 0.30-0.90.
- Rasch Item Difficulties should be within the -2.00 and +2.00 range; however, items difficulties around the mean of on-grade level tests are preferred.
- Choose items with high point biserial correlations; point biserial correlations should be greater than or equal to 0.30.
- Vertical linking items Rasch item fits should range from 0.80 to 1.20.