

## TELPAS Composite Reliability Estimates

The TELPAS composite scores are computed using student performance on the four language domains, where the domains are weighted using 5% *listening*, 5% *speaking*, 15% *writing*, and 75% *reading*. Because the *listening*, *speaking*, and *writing* domain scores for each student are ratings (ranging from 1 to 4) typically given by the student's English language teacher, the measurement errors for these three domains are assumed to be correlated. Reliability estimates of the TELPAS composite scores were calculated using a generalization of stratified  $\alpha$  method that allows for correlated measurement errors between the *listening*, *speaking*, and *writing* domains.

Two approaches were used to estimate the reliabilities of the TELPAS composite scores for all six grade clusters (2, 3, 4–5, 6–7, 8–9, 10–12) using the data collected in spring 2013 from all Texas students with limited English proficiency (LEP). These two approaches were: constrained estimation (i.e., constraining the *writing* domain reliability to the value obtained through an inter-rater reliability analysis conducted in 2008) and free estimation (i.e., estimating the *writing* domain reliability concurrently with the *listening*, *speaking*, and *writing* domains). For both approaches, the following steps were followed:

1. The reliability of the *reading* domain rating scores at each grade cluster was estimated using a method from Keng, Miller, O'Malley, & Turhan (2009).
2. The estimates of the reliabilities of the *listening*, *speaking*, and *writing* domain rating scores were computed using structural equation modeling (SEM), with the *writing* domain reliability value being either constrained or freely estimated.
3. The correlations among measurement errors for the *listening*, *speaking*, and *writing* domains were estimated as part of the SEM analyses.
4. The composite reliability estimate was computed for each grade cluster using the reliability estimates for the four domains.

Reliability estimates resulting from the analyses are presented in Tables 1 and 2. In sum, the reliability estimates for the TELPAS composites scores ranged from 0.89 to 0.92. Since internal consistency estimates 0.80 or greater are considered as adequate for group comparisons and estimates 0.90 and greater are considered adequate for individual applications (Nunnally & Bernstein, 1994), these 2012 estimates support reliable interpretations at the individual student level.

## References

- Keng, L., Miller, E., O'Malley, K.J., & Turhan, A. (2009). *A Generalization of Stratified  $\alpha$  that Allows for Correlated Measurement Errors between Subtests*. Retrieved March 1, 2013 from <http://www.pearsonassessments.com/NR/rdonlyres/8538688D-5827-4EB7-B63F-42D7742562BA/0/StratifiedAlphaThatAllowsforCorrelatedMeasurementErrorsbetweenSubtests.pdf>
- Nunnally, J., & Bernstein, I.H. (1994). *Psychometric theory (3rd ed.)*. New York: McGraw-Hill.

Table 1

## 2013 Estimated Reliability of TELPAS Composite Scores (Writing Freely Estimated)

Grade	Subject	$\mu$	$\sigma$	Internal consistency <sup>†</sup>	Reliability of composite
2 (n = 106058)	Listening	3.009	0.890	0.561	0.893
	Speaking	2.793	0.944	0.616	
	Writing	2.427	0.967	0.841	
	Reading <sup>††</sup>	2.858	1.009	0.852	
3 (n = 99546)	Listening	3.302	0.815	0.557	0.907
	Speaking	3.093	0.888	0.620	
	Writing	2.734	0.938	0.806	
	Reading	3.158	1.011	0.876	
4-5 (n = 153064)	Listening	3.487	0.754	0.544	0.894
	Speaking	3.309	0.832	0.614	
	Writing	3.020	0.895	0.810	
	Reading	3.300	0.919	0.859	
6-7 (n = 84263)	Listening	3.390	0.815	0.611	0.885
	Speaking	3.253	0.873	0.644	
	Writing	2.995	0.878	0.779	
	Reading	3.350	0.841	0.845	
8-9 (n = 59609)	Listening	3.248	0.925	0.726	0.922
	Speaking	3.101	0.986	0.745	
	Writing	2.930	0.938	0.805	
	Reading	3.282	0.971	0.894	
10-12 (n = 44018)	Listening	3.341	0.814	0.608	0.886
	Speaking	3.155	0.902	0.634	
	Writing	3.048	0.858	0.836	
	Reading	3.285	0.876	0.850	

Notes: <sup>†</sup>The internal consistency of Listening, Speaking, and Writing were estimated using SEM. The internal consistency of Reading on the categorical scale was estimated based on the internal consistency of Reading on the continuous scale.

<sup>††</sup>Only the online version of the Reading subtest was used.

Table 2

## 2013 Estimated Reliability of TELPAS Composite Scores (Writing Constrained)

Grade	Subject	$\mu$	$\sigma$	Internal consistency <sup>†</sup>	Reliability of composite
2 (n = 106058)	Listening	3.009	0.890	0.811	0.893
	Speaking	2.793	0.944	0.889	
	Writing	2.427	0.967	0.780	
	Reading <sup>††</sup>	2.858	1.009	0.852	
3 (n = 99546)	Listening	3.302	0.815	0.791	0.908
	Speaking	3.093	0.888	0.879	
	Writing	2.734	0.938	0.780	
	Reading	3.158	1.011	0.876	
4-5 (n = 153064)	Listening	3.487	0.754	0.779	0.895
	Speaking	3.309	0.832	0.880	
	Writing	3.020	0.895	0.780	
	Reading	3.300	0.919	0.859	
6-7 (n = 84263)	Listening	3.390	0.815	0.841	0.886
	Speaking	3.253	0.873	0.885	
	Writing	2.995	0.878	0.780	
	Reading	3.350	0.841	0.845	
8-9 (n = 59609)	Listening	3.248	0.925	0.879	0.922
	Speaking	3.101	0.986	0.901	
	Writing	2.930	0.938	0.780	
	Reading	3.282	0.971	0.894	
10-12 (n = 44018)	Listening	3.341	0.814	0.832	0.885
	Speaking	3.155	0.902	0.866	
	Writing	3.048	0.858	0.760	
	Reading	3.285	0.876	0.850	

Notes: <sup>†</sup>The inter-rater reliability for Writing was constrained using perfect agreement. The internal consistency of Reading on the categorical scale was estimated based on the internal consistency of Reading on the continuous scale. The internal consistency of Listening, and Speaking were estimated using SEM. <sup>††</sup>Only the online version of the Reading subtest was used.