

TELPAS Composite Reliability Estimates

The TELPAS composite scores are computed using student performance on the four language domains, where the domains are weighted using 10% *listening*, 10% *speaking*, 30% *writing*, and 50% *reading*. These domain weights were implemented for the first time in 2014 (see the 2013 TELPAS Composite Reliability Report for the previous set of domain weights). 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 α 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 2014 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 2013) 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). The reading proficiency level cut scores were reviewed and modified in the fall of 2013. These new reading standards were used for the 2014 TELPAS administration.
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 0.91 to 0.94. 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 2014 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 α 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

2014 Estimated Reliability of TELPAS Composite Scores (Writing Freely Estimated)

Grade	Subject	μ	σ	Internal consistency [†]	Reliability of composite
2 (n =108,164)	Listening	3.019	0.889	0.547	0.923
	Speaking	2.798	0.943	0.594	
	Writing	2.430	0.960	0.867	
	Reading ^{††}	2.444	0.990	0.852	
3 (n =102,222)	Listening	3.312	0.818	0.544	0.926
	Speaking	3.106	0.892	0.614	
	Writing	2.735	0.941	0.819	
	Reading	2.746	1.050	0.877	
4–5 (n =158,163)	Listening	3.502	0.757	0.526	0.916
	Speaking	3.323	0.838	0.600	
	Writing	3.024	0.902	0.834	
	Reading	2.785	0.923	0.849	
6–7 (n = 95,837)	Listening	3.444	0.803	0.581	0.911
	Speaking	3.312	0.864	0.614	
	Writing	3.049	0.882	0.826	
	Reading	2.673	0.890	0.843	
8–9 (n =67,331)	Listening	3.284	0.931	0.709	0.927
	Speaking	3.139	0.998	0.728	
	Writing	2.970	0.951	0.843	
	Reading	2.658	0.901	0.858	
10–12 (n =50,294)	Listening	3.368	0.805	0.615	0.912
	Speaking	3.186	0.897	0.643	
	Writing	3.074	0.854	0.832	
	Reading	2.732	0.895	0.847	

Notes: †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. ††Only the online version of the Reading subtest was used.

Table 2

2014 Estimated Reliability of TELPAS Composite Scores (Writing Constrained)

Grade	Subject	μ	σ	Internal consistency [†]	Reliability of composite
2 (n = 108,164)	Listening	3.019	0.889	0.815	0.936
	Speaking	2.798	0.943	0.885	
	Writing	2.430	0.960	0.909	
	Reading ^{††}	2.444	0.990	0.852	
3 (n = 102,222)	Listening	3.312	0.818	0.790	0.943
	Speaking	3.106	0.892	0.891	
	Writing	2.735	0.941	0.921	
	Reading	2.746	1.050	0.877	
4–5 (n = 158,163)	Listening	3.502	0.757	0.781	0.922
	Speaking	3.323	0.838	0.891	
	Writing	3.024	0.902	0.829	
	Reading	2.785	0.923	0.849	
6–7 (n = 95,837)	Listening	3.444	0.803	0.839	0.925
	Speaking	3.312	0.864	0.886	
	Writing	3.049	0.882	0.883	
	Reading	2.673	0.890	0.843	
8–9 (n = 67,331)	Listening	3.284	0.931	0.883	0.938
	Speaking	3.139	0.998	0.869	
	Writing	2.970	0.951	0.886	
	Reading	2.658	0.901	0.857	
10–12 (n = 50,294)	Listening	3.368	0.805	0.831	0.929
	Speaking	3.186	0.897	0.869	
	Writing	3.074	0.854	0.911	
	Reading	2.732	0.895	0.847	

Notes: †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. ††Only the online version of the Reading subtest was used.