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 *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 2011 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 2011 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. Under review.

Nunnally, J., Bernstein, I.H. Psychometric theory, 3rd ed. New York: McGraw-Hill; 1994.

Table 1

2011 Estimated Reliability of TELPAS Composite Scores (Writing Freely Estimated)

Grade	Subject	μ	σ	Internal consistency†	Reliability of composite
2 (n = 104303)	Listening	2.988	0.901	0.569	0.891
	Speaking	2.780	0.955	0.623	
	Writing	2.426	0.974	0.840	
	Reading††	2.919	0.963	0.847	
3 (n = 98188)	Listening	3.247	0.844	0.590	0.903
	Speaking	3.042	0.914	0.649	
	Writing	2.705	0.954	0.807	
	Reading	3.093	0.987	0.868	
4-5 (n = 138268)	Listening	3.398	0.800	0.593	0.901
	Speaking	3.211	0.873	0.656	
	Writing	2.953	0.922	0.796	
	Reading	3.350	0.877	0.865	
6-7 (n = 79291)	Listening	3.331	0.839	0.640	0.903
	Speaking	3.200	0.894	0.672	
	Writing	2.975	0.896	0.784	
	Reading	3.392	0.829	0.869	
8-9 (n = 57473)	Listening	3.139	0.943	0.714	0.923
	Speaking	2.982	1.003	0.728	
	Writing	2.833	0.957	0.828	
	Reading	3.220	0.956	0.895	
10-12 (n = 41612)	Listening	3.300	0.808	0.603	0.902
	Speaking	3.114	0.888	0.626	
	Writing	3.002	0.850	0.827	
	Reading	3.242	0.889	0.874	

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

2011 Estimated Reliability of TELPAS Composite Scores (Writing Constrained)

Grade	Subject	μ	σ	Internal consistency†	Reliability of composite
2 (n = 104303)	Listening	2.988	0.901	0.815	0.891
	Speaking	2.780	0.955	0.893	
	Writing	2.426	0.974	0.780	
	Reading††	2.919	0.963	0.847	
3 (n = 98188)	Listening	3.247	0.844	0.807	0.903
	Speaking	3.042	0.914	0.888	
	Writing	2.705	0.954	0.780	
	Reading	3.093	0.987	0.868	
4-5 (n = 138268)	Listening	3.398	0.800	0.798	0.902
	Speaking	3.211	0.873	0.883	
	Writing	2.953	0.922	0.780	
	Reading	3.350	0.877	0.865	
6-7 (n = 79291)	Listening	3.331	0.839	0.846	0.905
	Speaking	3.200	0.894	0.887	
	Writing	2.975	0.896	0.780	
	Reading	3.392	0.829	0.869	
8-9 (n = 57473)	Listening	3.139	0.943	0.875	0.923
	Speaking	2.982	1.003	0.894	
	Writing	2.833	0.957	0.780	
	Reading	3.220	0.956	0.895	
10-12 (n = 41612)	Listening	3.300	0.808	0.829	0.901
	Speaking	3.114	0.888	0.860	
	Writing	3.002	0.850	0.760	
	Reading	3.242	0.889	0.874	

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.