2016 TELPAS Composite Reliability Estimates

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 first implemented in 2014, and have remained the same through 2016. 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 (Keng, Miller, O'Malley, & Turhan, 2009) 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, and 10–12) using the data collected in spring 2016 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 during the 2016 TELPAS writing audit) 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 reliabilities of the *reading* domain rating scores for each grade cluster were estimated using a method from Keng, et al. (2009) to determine the reliability of the categorized rating score.
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, applying a Generalized Stratified α approach, 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 composite scores ranged from 0.91 to 0.94. Since internal consistency estimates of 0.80 or greater are considered as adequate for group comparisons and estimates of 0.90 and greater are considered adequate for individual applications (Nunnally & Bernstein, 1994), these 2016 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 October 21, 2016 from http://images.pearsonassessments.com/images/tmrs/tmrs_rg/StratifiedAlphathatAllowsforCorrelatedMeasurementErrorsbetweenSubtests.pdf?WT.mc_id=TMRS_A_Generalization_of_Stratified

Nunnally, J., & Bernstein, I.H. (1994). *Psychometric theory* (*3rd ed.*). New York: McGraw-Hill.

Table 1. 2016 Estimated Reliability of TELPAS Composite Scores *(Writing Freely Estimated)*

Grade	Subject	μ	σ	Internal consistency*	Reliability of composite
2 (n =111,297)	Listening	3.031	0.893	0.548	0.923
	Speaking	2.810	0.950	0.603	
	Writing	2.438	0.968	0.863	
	Reading	2.383	0.967	0.849	
3 (n =107,205)	Listening	3.305	0.829	0.539	0.923
	Speaking	3.096	0.899	0.610	
	Writing	2.705	0.945	0.829	
	Reading	2.749	1.009	0.866	
4–5 (n =176,614)	Listening	3.502	0.771	0.532	0.916
	Speaking	3.328	0.851	0.609	
	Writing	3.006	0.918	0.842	
	Reading	2.809	0.917	0.844	
6–7 (n = 119,061)	Listening	3.473	0.803	0.612	0.908
	Speaking	3.349	0.866	0.653	
	Writing	3.093	0.890	0.826	
	Reading	2.721	0.819	0.822	
8–9 (n =87,525)	Listening	3.297	0.932	0.707	0.922
	Speaking	3.160	0.995	0.732	
	Writing	2.991	0.953	0.856	
	Reading	2.560	0.865	0.836	
10–12 (n =65,757)	Listening	3.365	0.825	0.661	0.912
	Speaking	3.195	0.916	0.674	
	Writing	3.095	0.864	0.836	
	Reading	2.758	0.844	0.837	

*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.

Table 2. 2016 Estimated Reliability of TELPAS Composite Scores *(Writing Constrained)*

Grade	Subject	μ	σ	Internal consistency*	Reliability of composite
2 (n =111,297)	Listening	3.031	0.893	0.813	0.934
	Speaking	2.810	0.950	0.895	
	Writing	2.438	0.968	0.892	
	Reading	2.383	0.967	0.849	
3 (n =107,205)	Listening	3.305	0.829	0.791	0.942
	Speaking	3.096	0.899	0.895	
	Writing	2.705	0.945	0.935	
	Reading	2.749	1.009	0.866	
4–5 (n =176,614)	Listening	3.502	0.771	0.789	0.932
	Speaking	3.328	0.851	0.903	
	Writing	3.006	0.918	0.911	
	Reading	2.809	0.917	0.844	
6–7 (n = 119,061)	Listening	3.473	0.803	0.845	0.925
	Speaking	3.349	0.866	0.902	
	Writing	3.093	0.890	0.902	
	Reading	2.721	0.819	0.822	
8–9 (n =87,525)	Listening	3.297	0.932	0.883	0.937
	Speaking	3.160	0.995	0.914	
	Writing	2.991	0.953	0.921	
	Reading	2.560	0.865	0.836	
10–12 (n =65,757)	Listening	3.365	0.825	0.854	0.932
	Speaking	3.195	0.916	0.872	
	Writing	3.095	0.864	0.936	
	Reading	2.758	0.844	0.837	

*The internal consistency for Writing was constrained using the grade-band inter-rater reliability. 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.