

2009 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 the stratified *a* 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 2009 from all Texas students with limited English proficiency (LEP). These two approaches were: free estimation (i.e., estimating the *writing* domain reliability concurrently with the *listening*, *speaking*, and *writing* domains) and constrained estimation (i.e., constraining the *writing* domain reliability to the value obtained through an inter-rater reliability analysis conducted during the 2008 writing audit). 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 (2008).
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 0.90 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 2009 estimates support reliable interpretations at the individual student level.

References

- Keng, L., Miller, G.E., O'Malley, K. & Turhan, A. (2008). *A Generalization of Stratified a that Allows for Correlated Measurement Errors between Subtests*. Paper presented at the annual meeting of the American Educational Research Association, New York City, New York.
- Nunnally, J., Bernstein, I.H. *Psychometric theory*, 3rd ed. New York: McGraw-Hill; 1994.

Table 1. Estimated Reliabilities of the 2009 TELPAS Composite Scores
(Writing Freely Estimated)

Grade	Subject	μ	σ	Internal consistency [†]	Reliability of composite
2 (n = 100750)	Listening	2.869	0.945	0.591	0.896
	Speaking	2.667	0.995	0.634	
	Writing	2.343	1.004	0.852	
	Reading ^{††}	2.764	1.004	0.854	
3 (n = 93725)	Listening	3.084	0.905	0.621	0.906
	Speaking	2.896	0.960	0.664	
	Writing	2.583	0.982	0.808	
	Reading	3.013	1.013	0.871	
4-5 (n = 127094)	Listening	3.247	0.860	0.616	0.912
	Speaking	3.072	0.921	0.665	
	Writing	2.830	0.953	0.802	
	Reading	3.279	0.933	0.880	
6-7 (n = 72296)	Listening	3.177	0.886	0.642	0.915
	Speaking	3.042	0.935	0.667	
	Writing	2.838	0.931	0.795	
	Reading	3.300	0.897	0.886	
8-9 (n = 55693)	Listening	3.022	0.957	0.691	0.923
	Speaking	2.867	1.009	0.699	
	Writing	2.719	0.960	0.809	
	Reading	3.120	0.994	0.896	
10-12 (n = 43725)	Listening	3.211	0.822	0.579	0.898
	Speaking	3.042	0.890	0.595	
	Writing	2.923	0.855	0.809	
	Reading	3.263	0.883	0.870	

Notes: [†]The internal consistency reliabilities of the *listening*, *speaking*, and *writing* domains were estimated using a structural equation modeling approach.

^{††}The internal consistency of *reading* rating scores was estimated based on the internal consistency of *reading* raw scores. The 2009 online data were used in the analysis.

Table 2. Estimated Reliabilities of the 2009 TELPAS Composite Scores
(Writing Constrained using the 2008 Writing Audit Reliability Estimates)

Grade	Subject	μ	σ	Internal consistency [†]	Reliability of composite
2 (n = 100750)	Listening	2.869	0.945	0.833	0.896
	Speaking	2.667	0.995	0.894	
	Writing	2.343	1.004	0.780	
	Reading ^{††}	2.764	1.004	0.854	
3 (n = 93725)	Listening	3.084	0.905	0.829	0.906
	Speaking	2.896	0.960	0.886	
	Writing	2.583	0.982	0.780	
	Reading	3.013	1.013	0.871	
4-5 (n = 127094)	Listening	3.247	0.860	0.821	0.913
	Speaking	3.072	0.921	0.886	
	Writing	2.830	0.953	0.780	
	Reading	3.279	0.933	0.880	
6-7 (n = 72296)	Listening	3.177	0.886	0.854	0.916
	Speaking	3.042	0.935	0.886	
	Writing	2.838	0.931	0.780	
	Reading	3.300	0.897	0.886	
8-9 (n = 55693)	Listening	3.022	0.957	0.880	0.924
	Speaking	2.867	1.009	0.890	
	Writing	2.719	0.960	0.780	
	Reading	3.120	0.994	0.896	
10-12 (n = 43725)	Listening	3.211	0.822	0.829	0.898
	Speaking	3.042	0.890	0.852	
	Writing	2.923	0.855	0.760	
	Reading	3.263	0.883	0.870	

Notes: [†]The internal consistency reliabilities of the *listening* and *speaking* domains were estimated using a structural equation modeling approach.

^{††}The internal consistency of *reading* rating scores was estimated based on the internal consistency of *reading* raw scores. The 2009 online data were used in the analysis.