

2010 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 2010 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 2010 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 2010 TELPAS Composite Scores
(Writing Freely Estimated)

Grade	Subject	μ	σ	Internal consistency [†]	Reliability of composite
2 (n = 102202)	Listening	2.922	0.933	0.585	0.897
	Speaking	2.711	0.985	0.631	
	Writing	2.368	0.997	0.838	
	Reading ^{††}	2.846	1.014	0.856	
3 (n = 96651)	Listening	3.135	0.875	0.599	0.907
	Speaking	2.936	0.941	0.651	
	Writing	2.610	0.964	0.811	
	Reading	3.123	0.996	0.873	
4-5 (n = 131687)	Listening	3.299	0.833	0.596	0.902
	Speaking	3.110	0.901	0.649	
	Writing	2.849	0.931	0.797	
	Reading	3.281	0.907	0.867	
6-7 (n = 75299)	Listening	3.260	0.867	0.629	0.903
	Speaking	3.134	0.913	0.656	
	Writing	2.895	0.912	0.789	
	Reading	3.297	0.862	0.868	
8-9 (n = 56824)	Listening	3.083	0.951	0.702	0.924
	Speaking	2.931	1.010	0.715	
	Writing	2.792	0.963	0.829	
	Reading	3.158	0.997	0.897	
10-12 (n = 42674)	Listening	3.249	0.816	0.591	0.895
	Speaking	3.075	0.887	0.608	
	Writing	2.967	0.851	0.818	
	Reading	3.282	0.843	0.864	

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 2010 online data were used in the analysis.

Table 2. Estimated Reliabilities of the 2010 TELPAS Composite Scores
(Writing Constrained using the 2008 Writing Audit Reliability Estimates)

Grade	Subject	μ	σ	Internal consistency [†]	Reliability of composite
2 (n = 102202)	Listening	2.922	0.933	0.824	0.897
	Speaking	2.711	0.985	0.889	
	Writing	2.368	0.997	0.780	
	Reading ^{††}	2.846	1.014	0.856	
3 (n = 96651)	Listening	3.135	0.875	0.815	0.907
	Speaking	2.936	0.941	0.886	
	Writing	2.610	0.964	0.780	
	Reading	3.123	0.996	0.873	
4-5 (n = 131687)	Listening	3.299	0.833	0.807	0.903
	Speaking	3.110	0.901	0.879	
	Writing	2.849	0.931	0.780	
	Reading	3.281	0.907	0.867	
6-7 (n = 75299)	Listening	3.260	0.867	0.847	0.904
	Speaking	3.134	0.913	0.884	
	Writing	2.895	0.912	0.780	
	Reading	3.297	0.862	0.868	
8-9 (n = 56824)	Listening	3.083	0.951	0.880	0.924
	Speaking	2.931	1.010	0.896	
	Writing	2.792	0.963	0.780	
	Reading	3.158	0.997	0.897	
10-12 (n = 42674)	Listening	3.249	0.816	0.834	0.895
	Speaking	3.075	0.887	0.857	
	Writing	2.967	0.851	0.760	
	Reading	3.282	0.843	0.864	

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 2010 online data were used in the analysis.