

## Reliability and Validity Study of the edTPA @ SCSU

### Summary

1. The first sample of scores from the *EdTPA* suggests that the scales behave in an internally consistent manner, suggesting that they are reliable.
2. The conclusion above may be mitigated somewhat by the finding that all 13 items (rubrics) proved internally consistent ( $\alpha = .92$ ). This raises the possibility that the reliability of scales (domains/areas) may exist as an artifact of the internal consistency of the entire instrument. We will continue to monitor this possibility. Eventually, enough cases will accrue that structural analyses can be undertaken.
3. As suggested by the reliability data, all correlations between *EdTPA* domains proved significant, ranging from .81 (“Planning” and “Academic Language”) to .32 ( $p = .014$ ) (“Analyzing Instruction” and “Instruction”).
4. It would be useful to see the internal consistency reliabilities for scales rise to near or over .90 before the *EdTPA* is employed as a high-stakes instrument for individual candidates. We will continue to monitor this situation.
5. We will have to rely on analysis of the national sample for inter-rater agreement levels for individual items. In addition, no internal consistency value for “Analysis” can be calculated at present because it consists of one item. At this level of analysis, no reliability evaluation can be conducted for individual items (e.g., rubrics).
6. Perhaps the best news is that positive discriminant validity data exists on a very preliminary basis. Specifically, no significant correlations were observed between *Basic Skills—Writing* and *EdTPA*. To the extent that one considers *MTLE-Writing* reliable and valid, this suggests that scorers have successfully distinguished between the skills assessed via the *EdTPA* and levels of written expression ability. Of course, this is based on a small number of cases and deserves further monitoring, especially for the case of the *EdTPA* “Assessment” domain.

### Context/ Background

This analysis is based on the first 60 cases returned from the trained scorers via the Pearson system. We added 20 cases where candidates completing the Teacher Performance Assessment (Hereafter *EdTPA*) had also completed the *MTLE Basic Skills Writing*. While several candidates, 7/20, 35 %, had taken *MTLE Writing* more than one time, we included only their highest score in the analysis. For the  $N = 2$  cases where some *EdTPA* scores were missing, mean scores were inserted to increase  $N$ .

We had 60 interpretable cases; the mean scores by area are shown below. In addition, Chronbach’s coefficient alpha was calculated for each domain including more than one item. Coefficient  $\alpha$  reflects the degree of internal consistency within scales, as it nears a value of one, items appear to be reliably working together to measure an entity. The right-most column consists of the Pearson product-moment correlation between *EdTPA* vales and those for *MTLE Writing*.

Table 1. Mean EdTPA scores (and reliability estimates) by domain: Sixty initial cases.

Area	N	Mean	SD	Reliability (Cronbach's $\alpha$ )	$r_{xy}$ with <i>MTLE Writing</i> ( $N = 20$ )
Planning (4 items)	60	2.9	.75	.85	.01 <sup>1</sup>
Instruction (2 items)	60	2.7	.73	.89	.16
Assessment (3 items)	60	2.3	.78	.82	.28 <sup>2</sup>
Academic Language (3 items)	60	2.6	.80	.79	-.02
Analysis of teaching (1 item)	60	2.5	.72	----	-.15
TOTAL EdTPA (13 items)	60	2.6	.61	.92	.07

<sup>1</sup>None of the bivariate correlations between *EDTPA* and *MTLE Writing* were statistically significant

<sup>2</sup>Nonsignificant, but  $p$  of .23 suggests that this ought to be monitored on an ongoing basis.