Candidate Preservice Assessment of Student Teaching (CPAST) Form Summary

What is the CPAST Form?

A **formative** and **summative assessment** during the **student teaching practicum**.

- The 21-row rubric has **two subscales:** (1) Pedagogy and (2) Dispositions with detailed descriptors of observable, measurable behaviors, to guide scoring decisions.
- An additional "Look Fors" resource provides and elaborates on the qualities and behaviors for a given level of performance (i.e., evidence and sources of evidence).
- A self-paced **90-minute training module** is available for users of the Form.

What analyses did we perform on the Form data?

We explored:

- Validity (content, construct and concurrent)
- Reliability (internal consistency, inter-rater reliability)

Who were the participants?

- During the academic year of 2015-2016 we collected valid data from **1203 teacher candidates** from **23 EPPs in Ohio**.
- Of the 1203 teacher candidates, **32** were recruited to participate in the inter-rater reliability study, in which each teacher candidate was evaluated by two supervisors – their primary university supervisor (i.e., the supervisor who was formally assigned by the EPPs to supervise the teacher candidate during the student teaching), and a secondary rater (i.e., a supervisor who completed a minimum of three observations of the teacher candidates throughout the semester).

What were the findings?

Validity and reliability met standards for instrument development. Below is a short description of evidences of validity and reliability of the instrument. More detailed analysis can be obtained upon request.

Content Validity

- Investigated by calculating a **content validity ratio** (CVR; Lawshe, 1975) for the aspects of clarity, importance, and representativeness of the CPAST Form. [CVR=$\frac{n_e-(N/2)}{N/2}$, where E refers to the number of experts who rated the item as equal to or above 3, and N refers to the total number of experts].
- Three experts (a K-12 teacher, a university teacher education professor, and a psychometrician) provided ratings of these aspects on a scale of one to four.
- Clarity: All items (except Row D in Pedagogy and Row G in Disposition), reached a CVR of 1. **The average CVR for all the items was 0.94**, exceeding the criterion of 0.8, indicating that the scale had strong content validity for clarity.
- Importance: **All items reached a value of 1**, revealing that all the item questions were important in measuring the constructs of pedagogy and disposition.
- Representativeness: All items (except Row H in Pedagogy and Row G in Disposition) reached a value of 1. **The average CVR for all the items was 0.94**, suggesting that the rows were representative of the theoretical domain of the constructs.

Construct Validity

- Confirmatory factor analysis (CFA) was conducted using Mplus Version 7.11 (Muthén & Muthén, 1998-2015) to examine the construct validity.
- The estimator of weighted least squares with mean and variance adjustment (WLSMV) was adopted, which was demonstrated to be suitable for handling ordinal data (Flora & Curran, 2004).
- The three indices selected for this study were the root mean-square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker–Lewis index (TLI), and the model fit was evaluated based on the following criteria: RMSEA <.06, CFI >.95, and TLI >.95 (Hu & Bentler, 1999).
- The **model fit indexes** RMSEA (0.048), CFI (0.980) and TLI (0.978) indicated that the **hypothesized two-factor model fit the data reasonably well**; the loadings ranged from 0.676 to 0.841, all at .001 significance level, indicating that all the items are moderately or strongly associated with their corresponding latent factors. Figure 1 (p. 4) displays the two-factor model of CPAST Form.
- The **Pedagogy and Dispositions scales were highly correlated** (r= .873, p <.001), indicating a strong association between a teacher candidate's teaching knowledge/skills and dispositions.
- The **correlation between the two latent factors was in concordance with existent literature**, which supports that teachers' professional dispositions and teaching practice are closely linked to each other (Kuzborska, 2011).

Longitudinal Measurement Invariance

- Longitudinal invariance was tested through a hierarchy of nested models. In Table 1, Model 1, Model 2, and Model 3 refer to the configural invariance model, weak factorial invariance model and strong factorial invariance model.
- The **configural invariance model had good model fit** (RMSEA = 0.051, CFI = 0.978, TLI = 0.976). The **weak factorial invariance model also had good fit** (RMSEA = 0.040, CFI = 0.986, TLI = 0.985). Additionally, **the weak factorial invariance model did not fit worse compared to the configural invariance model** ($\Delta\chi2$ = 17.658, Δdf = 19, p = .5454), and **all the differences in terms of CFI, TLI, and RMSEA were close to or less than .01.** The **strong factorial invariance model did fit worse compared to the weak factorial invariance model** ($\Delta\chi^2$ = 158.257, Δdf = 40, p=.0000).
- The results suggest that **the instrument has weak factorial invariance, suggesting the same latent variances are being measured across time.**

Table 1. Longitudinal Measurement Invariance

Models	χ^2	df	RMSEA	CFI	TLI	$\Delta\chi^2$	Δdf	p	ΔRMSEA	ΔCFI	ΔTLI
Model 1	1541.134	376	0.051	0.978	0.976	--	--	--	--	--	--
Model 2	1154.712	395	0.040	0.986	0.985	17.658	19	0.5454	-0.011	0.008	0.009
Model 3	1285.544	435	0.040	0.984	0.985	158.257	40	0.0000	0.000	-0.002	0.000
Model 4	1194.985	426	0.039	0.986	0.986	43.964	31	0.0614	-0.001	0.000	0.001

Note: Model 1= configural factorial invariance model
Model 2= weak factorial invariance model
Model 3= strong factorial invariance model
Model 4= partial strong factorial invariance model

Inter-rater Reliability

- Table 2 reports two reliability statistics: **adjacent agreement** and **Kappa-n**. Adjacent agreement refers to the proportion of cases in which two independent scorers assign either the exact same

score or a score within 1 point of each other. When scoring complex performance assessment tasks, this approach is often used as a measure of rater agreement. In some cases, scorers will assign the same score simply by chance. Kappa-n κ_n adjusts the adjacent agreement rate to take into account this chance agreement.

- The average **adjacent agreement rate was 98%** and the **average Kappa-n was 0.97**.
- Although several types of reliability analyses were conducted to examine agreement rates between scorers on the CPAST Form, **these two statistics were reported here because SCALE (2013) used them when assessing the inter-rater reliability of edTPA**.

Table 2 *Rubric Row Inter-rater Reliability*

Item	Agreement Rate	Kappa-N
Focus for Learning: Standards and Objectives/Targets	100%	1.00
Materials and Resources	100%	1.00
Assessment of P-12 Learning	100%	1.00
Differentiated Methods	100%	1.00
Learning Target and Directions	100%	1.00
Critical Thinking	100%	1.00
Checking for Understanding and Adjusting Instruction through Formative Assessment	100%	1.00
Digital Tools and Resources	100%	1.00
Safe and Respectful Learning Environment	96.9%	0.96
Data-Guided Instruction	100%	1.00
Feedback to Learners	100%	1.00
Assessment Techniques	100%	1.00
Connections to Research and Theory	100%	1.00
Participates in Professional Development	87.5%	0.83
Demonstrates Effective Communication with Parents or Legal Guardians	87.5%	0.85
Demonstrates Punctuality	90.6%	0.86
Meets Deadlines and Obligations	100%	1.00
Preparation	96.9%	0.96
Collaboration	96.9%	0.96
Advocacy to Meet the Needs of Learners or for the Teaching Profession	96.9%	0.96
Responds Positively to Constructive Criticism	96.9%	0.96

Internal consistency reliability

- Examined by calculating the Cronbach Alpha coefficient using SPSS statistical package version 23.0.
- Results show the Cronbach's Alpha coefficient is **0.907 for the Pedagogy subscale**, **0.831 for the Dispositions subscale**, and **0.929 for the total scale**, suggesting that the subscales and the total scale **display good internal consistency**.