Wow – There’s a Valid and Reliable Student Teaching Instrument!

Xiangquan (James) Yao, Erica Brownstein, and Carolyn Kaplan (The Ohio State Univ.)
Kristall J. Day (Ohio Dominican Univ.)
Agenda
Candidate Preservice Assessment of Student Teaching (CPAST)

- CPAST Form
  - **Who** designed
  - **How** designed
  - Instrument description
    - Training
    - Resources

- Evidence of Validity & Reliability
  - Data collection
  - Demographic data
  - Analyses conducted
  - Results

- Example CPAST Data for Participating Institutions
Pre-Assessment

• Does your EPP use a unit-wide instrument to assess student teaching?

• Do you train your supervisors to use your instrument?

• Has the instrument been analyzed for validity and reliability?
Discuss with your neighbor

• What are three (3) essential components in a formative and summative student teaching assessment form?
It takes an army to build a form...

The Ohio State University
  F. Beickelman
  M. Bendixen-Noe
  P. Bode
  E. Brownstein
  K. Day
  M. Fresch
  C. Kaplan
  C. Warner
  M. Whittington

Wilmington College
  M. Hendricks

University of Toledo
  V. Stewart

University of Akron
  W. Jewell

Cleveland State University
  A. Price
  A. Crell

Ohio University
  C. Patterson

Kent State University
  J. Arhar
  S. Turner

Wright State University
  T. Kahrig

Bowling Green State University
  D. Gallagher

Wittenberg University
  S. Brannan
  T. Whitlock
<table>
<thead>
<tr>
<th>The Ohio State Univ.</th>
<th>Univ. of Dayton</th>
<th>Kent State Univ.</th>
<th>Wilmington College</th>
<th>Wright State Univ.</th>
<th>Bowling Green State Univ.</th>
<th>Univ. of Toledo</th>
<th>Univ. of Akron</th>
<th>Cleveland State Univ.</th>
<th>Ohio Univ.</th>
<th>Wittenberg Univ.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univ. of Findlay</td>
<td>Baldwin Wallace Univ.</td>
<td>Cedarville Univ.</td>
<td>Cincinnati Christian Univ.</td>
<td>Lake Erie College</td>
<td>Lourdes Univ.</td>
<td>Malone Univ.</td>
<td>Marietta College</td>
<td>Notre Dame College</td>
<td>John Carroll Univ.</td>
<td>Walsh Univ.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**VARI-EPP Project Collaboration**
What is the CPAST Form?

Candidate Preservice Assessment of Student Teaching (CPAST) Form

• A formative and summative assessment during student teaching, aligned to CAEP and InTASC Standards.

• The rubric has two subscales:
  • (1) Pedagogy and
  • (2) Dispositions

• Developed over three years
Each row includes detailed descriptors of observable, measurable behaviors to guide scoring decisions.

<table>
<thead>
<tr>
<th>Item</th>
<th>Exceeds Expectations (3 points)</th>
<th>Meets Expectations (2 points)</th>
<th>Emerging (1 point)</th>
<th>Does not meet Expectation (0 points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. Data-Guided Instruction</td>
<td>Uses <strong>data-informed decisions</strong> <em>(trends and patterns)</em> to set short and long term goals for future instruction and assessment AND Uses <strong>contemporary tools</strong> for learner <strong>data</strong> record-keeping and <strong>analysis</strong></td>
<td>Uses <strong>data-informed decisions</strong> to design instruction and assessment AND Uses <strong>contemporary tools</strong> for learner <strong>data</strong> record-keeping</td>
<td>Uses minimal <strong>data</strong> to design instruction and assessment</td>
<td>Does not use <strong>data</strong> to design instruction and assessment</td>
</tr>
</tbody>
</table>
An additional “Look Fors” resource elaborates on the qualities and behaviors for a given level of performance, describing:

- Sources of Evidence

### H. Digital Tools and Resources

<table>
<thead>
<tr>
<th>Item</th>
<th>Exceeds Expectations (5 points)</th>
<th>Meets Expectations (2 points)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discusses AND uses a variety of developmentally appropriate technologies (digital tools and resources) that</td>
<td>Discusses AND uses developmentally appropriate technologies (digital tools and resources) that</td>
</tr>
<tr>
<td></td>
<td>1. Are relevant to learning objectives/targets of the lesson</td>
<td>1. Are relevant to learning objectives/targets of the lesson</td>
</tr>
<tr>
<td></td>
<td>2. Engage learners in the demonstration of knowledge or skills</td>
<td>2. Engage learners in the demonstration of knowledge or skills</td>
</tr>
<tr>
<td></td>
<td>3. Extend learners’ understanding of concepts</td>
<td>3. Extend learners’ understanding of concepts</td>
</tr>
</tbody>
</table>

### Sources of Evidence:

- Observation of teaching (Refer to VARI-EPP Student Teaching Form Glossary)
- Pre/post observation conferences
- Cumulative lesson plans
- Conversations with and/or documentation from the mentor teacher

### Possible Evidence:

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Exceeds/Meets Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student teacher uses and discusses some of the following digital tools:</td>
<td></td>
</tr>
<tr>
<td>Computers</td>
<td></td>
</tr>
<tr>
<td>Websites</td>
<td></td>
</tr>
<tr>
<td>Blogs</td>
<td></td>
</tr>
<tr>
<td>Mobile devices</td>
<td></td>
</tr>
<tr>
<td>Interactive whiteboards</td>
<td></td>
</tr>
<tr>
<td>Online media</td>
<td></td>
</tr>
<tr>
<td>Online study tools</td>
<td></td>
</tr>
<tr>
<td>Student teacher uses digital tools in the following ways:</td>
<td></td>
</tr>
<tr>
<td>Relevant - Directly support access to the objectives for the lesson(s)</td>
<td></td>
</tr>
<tr>
<td>Engaging - Learners are actively using the digital tools instead of the teacher just using the tools and learners are passive</td>
<td></td>
</tr>
<tr>
<td>Extending - Learners are given independent assignments to use digital tools to continue exploring a topic</td>
<td></td>
</tr>
</tbody>
</table>

* This row received low Inter-Rater Reliability scores in the first round of data collection.

Updated 8/1/16 © 2016
CPAST Form Training for Supervisors

• Parameters
  • Required & Sustainable
  • Assessments for both
    • 80% correct required to pass

• Training
  • Initial version
    • Three chapters
    • ~90 minutes to complete (self-paced online)
  • Refresher(s) version
    • One chapter
    • ~30 minutes to complete (self-paced online)
CASE STUDY:
You are Carry Candidate’s supervisor.

- She is completing her student teaching in a building with technology limited to an overhead and a projected computer screen.
- Although the school has limited resources, students do have cell phones with capabilities to text.

In your observations of her teaching in the first half of the semester, she used PowerPoint to present a lesson, and the content of the slides was aligned to her learning targets.

- To share her lesson plans and materials with you and her cooperating teacher, Carry routinely places her Word documents and PowerPoint presentations in the DropBox file sharing site.
- Additionally, during her reflections she has indicated what she would do if she had advanced technology in her classroom.

Based on the above evidence, how would you score Carry at the midterm evaluation? (Select one response)

A) Exceeds Expectations  B) Meets Expectations
C) Emerging  D) Does Not Meet Expectations
Does Not Meet Expectations (0 points)

One of the following:
A. Does not use technologies to engage learners (digital tools and resources)
AND
Technology is available in the setting
OR
B. Use of technologies is not relevant to the learning objectives/targets of the lesson
OR
C. Does not discuss technologies
AND
Technology is not available in the setting

Yes, Carry uses PowerPoint, but there was no evidence her lessons engaged learners.

Technology that could engage students, even though limited, is available through computer screen projection and student cell phones.

From the Look Fors

Emerging/Does Not Meet Expectations
- Student teacher uses technology "on stage" with little learner interaction (ISTE Essential Conditions Rubric)
- Student teacher "uses technology for own productivity in relationship to teaching and learning" (ISTE Essential Conditions Rubric)

Carry did use technology, but there was little interaction by the P-12 learners.

Carry also used technology for her own productivity. Again, this is not sufficient to "meet expectations" for the row.
From the Look Fors

**Emerging/ Does Not Meet Expectations**

- Student teacher uses technology “on stage” with little learner interaction (ISTE Essential Conditions Rubric)
- Student teacher “uses technology for own productivity in relationship to teaching and learning” (ISTE Essential Conditions Rubric)

Carry did use technology, but there was **little interaction** by the P-12 learners.

Carry also used technology for her own productivity. Again, this is **not sufficient** to “meet expectations” for the row.

Yes, Carry uses PowerPoint, but there was **no evidence** her lessons engaged learners.

Technology that could engage students, even though limited, is available through computer screen projection and student cell phones.
Congratulations!

You completed training for one row!
Confer with Your Buddy

• Is this training a task *your supervisors* could accomplish?
• Do you see a need for this type of training?
• What strategies might you use to ensure your supervisors completed a two-hour training requirement?

• Remember
  • Initial training only has three chapters & takes ~90 minutes
  • Refreshers are only one chapter and take ~30 minutes
Additional Resource: Cooperating Teacher/Student Teacher Training

- Optional
- Contains
  - One chapter (i.e., PowerPoint presentation)
  - No assessments
- ~20 minutes to complete
Evidence of Validity and Reliability

What Data Were Collected?

**Data**
- Midterm & Final CPAST Consensus Score
  - Combined score from Supervisor, Candidate, & Student Teacher
- edTPA scores
- State test scores
- Demographics

**All IHEs**

**IRR IHEs**

Data plus
- three observations
- IRR Supervisor final evaluation
### Demographic Data of Participating Institutions

<table>
<thead>
<tr>
<th>Institution Type</th>
<th># of Institutions</th>
<th># of Student Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban - Public</td>
<td>6</td>
<td>481</td>
</tr>
<tr>
<td>Suburban - Public</td>
<td>2</td>
<td>359</td>
</tr>
<tr>
<td>Rural - Public</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Urban - Private</td>
<td>5</td>
<td>208</td>
</tr>
<tr>
<td>Suburban - Private</td>
<td>6</td>
<td>85</td>
</tr>
<tr>
<td>Rural - Private</td>
<td>4</td>
<td>70</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>23</strong></td>
<td><strong>1203</strong></td>
</tr>
</tbody>
</table>
Demographic Data of Participating Student Teachers

<table>
<thead>
<tr>
<th>Placement Type</th>
<th># of Student Teachers</th>
<th>% of Student Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>447</td>
<td>37.2%</td>
</tr>
<tr>
<td>Suburban</td>
<td>542</td>
<td>45.1%</td>
</tr>
<tr>
<td>Rural</td>
<td>299</td>
<td>24.9%</td>
</tr>
<tr>
<td>Missing</td>
<td>95</td>
<td>7.9%</td>
</tr>
</tbody>
</table>

* N = 1383 due to multiple placements

<table>
<thead>
<tr>
<th>License Area</th>
<th># of Student Teachers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Childhood Education</td>
<td>465</td>
<td>38.7%</td>
</tr>
<tr>
<td>Middle Childhood Education</td>
<td>202</td>
<td>16.8%</td>
</tr>
<tr>
<td>Special Education</td>
<td>168</td>
<td>14.0%</td>
</tr>
<tr>
<td>Secondary English Lang. Art</td>
<td>99</td>
<td>8.2%</td>
</tr>
<tr>
<td>Social Studies Education</td>
<td>92</td>
<td>7.6%</td>
</tr>
<tr>
<td>Secondary Math Education</td>
<td>70</td>
<td>5.8%</td>
</tr>
<tr>
<td>Secondary Science Education</td>
<td>50</td>
<td>4.2%</td>
</tr>
<tr>
<td>Music Education</td>
<td>39</td>
<td>3.2%</td>
</tr>
<tr>
<td>Foreign Language</td>
<td>29</td>
<td>2.4%</td>
</tr>
<tr>
<td>Physical Education</td>
<td>14</td>
<td>1.2%</td>
</tr>
<tr>
<td>Art Education</td>
<td>2</td>
<td>0.2%</td>
</tr>
<tr>
<td>TESOL</td>
<td>1</td>
<td>0.1%</td>
</tr>
<tr>
<td>Family and Consumer Science Education</td>
<td>1</td>
<td>0.1%</td>
</tr>
<tr>
<td>Undisclosed</td>
<td>4</td>
<td>0.3%</td>
</tr>
</tbody>
</table>
Demographic Data of Participating Student Teachers

P-12 School Level

- **Elementary**: 697 (58%)
- **Middle**: 300 (25%)
- **High School**: 379 (32%)

*N = 1376 due to multiple placements.*
Comparison: Demographic Data of CPAST Student Teachers & National Teacher Statistics

Comparison % CPAST Student Teachers & Practicing Teachers in US by Race

(Feistritzer, Grffing, & Linnajarvi, 2011)
Comparison: Demographic Data of CPAST Student Teachers & National Teacher Statistics

Comparison % of CPAST Student Teachers & Practicing Teachers in the US by Gender

(Feistritzer, Grffing, & Linnajarvi, 2011)
Types of Validity and Reliability Analysis Conducted

Validity (AERA, 2014)
- Content Validity
- Construct Validity
- Concurrent Validity

Reliability (AERA, 2014)
- Test-retest Reliability
- Internal Consistency Reliability
- Interrater Reliability
- Addressed areas in the CAEP Evaluation Framework for Assessments
## CAEP Evaluation Framework for Assessments

### Examples of Attributes Below Sufficient Level

- Description of or plan to establish validity does not inform reviewers about how it was established or is being investigated.
- The type of validity established or investigated is miss-identified or not described.
- The instrument was not piloted before administration.
- Process or plans for data analysis and interpretation are not presented or are superficial.
- Described steps do not meet accepted research standards for establishing validity. For example, validity is determined through an internal review by only one or two stakeholders.

### CAEP Sufficient Level

#### DATA VALIDITY

- A description or plan is provided that details steps the EPP has taken or is taking to ensure the validity of the assessment and its use.
- The plan details the types of validity that are under investigation or have been established (e.g., construct, content, concurrent, predictive, etc.) and how they were established.
- If the assessment is new or revised, a pilot was conducted.
- The EPP details its current process or plans for analyzing and interpreting results from the assessment.
- The described steps meet accepted research standards for establishing the validity of data from an assessment.

### Examples of Attributes Above Sufficient Level

- Types of validity investigated go beyond content validity and move toward predictive validity.
- A validity coefficient is reported.

---

**Statistical Analyses**

**NOT doing predictive validity at present**

*Contact if interested!* 😊

Survey of Experts
### CAEP Evaluation Framework for Assessments

#### 4. DATA RELIABILITY

<table>
<thead>
<tr>
<th>EXAMPLES OF ATTRIBUTES BELOW SUFFICIENT LEVEL</th>
<th>CAEP SUFFICIENT LEVEL</th>
<th>EXAMPLES OF ATTRIBUTES ABOVE SUFFICIENT LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Description of or plan to establish reliability does not inform reviewers about how it was established or is being investigated.</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>b. Described steps do not meet accepted research standards for reliability.</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>c. No evidence, or limited evidence, is provided that scorers are trained, and their inter-rater agreement is documented.</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>d. Described steps do not meet accepted research standards for reliability.</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

- **Statistical Analyses:** midterm compared to final
- **Statistical Analyses:** Comparison for agreement between two independent raters

**CPAST Validity & Reliability**
Content Validity

• Method: Content Validity Ratio
• Participants: Three Experts
  • K-12 teacher
  • EPP Faculty
  • Psychometrician
• Measurement & Results
  • Clarity = average CVR of 0.94
  • Importance = all items reached a value of 1
  • Representativeness = average CVR of 0.94

• The results indicate CPAST has a high content validity (Wilson, Pan & Schumsky, 2012)
Construct Validity: CFA Results

- **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.

- Factor loadings range 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.

- Pedagogy and Dispositions scales were correlated (r=0.873), which is supported by the literature (Kuzborska, 2011).
Construct Validity: CFA Results

- **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.

- Factor loadings range 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.

- Pedagogy and Dispositions scales were correlated (r=0.873), which is supported by the literature (Kuzborska, 2011).
Two-Factor Model of CPAST
Reliability: Test-retest & Internal Consistency

Test-retest Reliability

• Pedagogy & Disposition \textit{significantly correlated} at midterm and final term
• (Pedagogy = .752; Disposition = .845)

Internal consistency reliability

• Cronbach Alpha coefficient Results
  • 0.907 for the Pedagogy subscale
  • 0.831 for the Dispositions subscale
  • 0.929 for the total scale
• Subscales and the total scale display \textit{good internal consistency}

*Note: A commonly accepted rule of thumb for describing internal consistency is as follows: \( \alpha \geq 0.9 \) Excellent, \( 0.9 > \alpha \geq 0.8 \) Good, \( 0.8 > \alpha \geq 0.7 \) Acceptable, \( 0.7 > \alpha \geq 0.6 \) Questionable, \( 0.6 > \alpha \geq 0.5 \) Poor, \( 0.5 > \alpha \) Unacceptable
Inter-rater Reliability

- Adjacent agreement and Kappa-n statistics
  - Same methods used by SCALE for edTPA analyses

- Adjacent agreement
  - Same or within one

- Kappa-n
  - Accounts for chance agreements

- Exceeds CAEP requirements
  - Adjacent agreement = 98%
  - Average Kappa-n = 0.97

*Note: CAEP requires assessment to have an interrater reliability greater than 0.8.
Example CPAST Reports for Institutions

- Do you have sufficient **N** to conduct the previous V&R studies?
- Do you have the **resources** to conduct the previous V&R studies?
- Do you have the **infrastructure** to conduct the previous V&R studies?
- Do you have the **desire** (😊) to conduct the previous V&R studies?
- Do you have **comparison** scores (with other EPPs)?
Example CPAST Reports for Institutions
CPAST Reports for Institutions

Each participating institution will receive these data reports every semester.

**Institutional Data Report**
Scores given
- Overall
- Gender
- Race
- Degree level
- Each program area
- Placement setting

**Comparison Data Report**
Scores given
- Overall
- Gender
- Race
- Degree level
- Each program area
- Placement setting
Mean by Row

Sample Data Provided to Participating Institutions

Overall Mean by Subscale
Would YOU like to join our team?

What questions do you have?
Thank you for your time!

- If you have any questions AT ANY TIME, feel free to contact:

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
<th>Phone</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erica Brownstein</td>
<td><a href="mailto:Brownstein.2@osu.edu">Brownstein.2@osu.edu</a></td>
<td>(614) 292-1414</td>
<td>“Big Picture” project questions, rubric questions</td>
</tr>
<tr>
<td>Carolyn Kaplan</td>
<td><a href="mailto:Kaplan.169@osu.edu">Kaplan.169@osu.edu</a></td>
<td>(614) 292-5044</td>
<td>Training, Timeline, IRR Participants</td>
</tr>
<tr>
<td>Xiangquan (James) Yao</td>
<td><a href="mailto:Yao.298@osu.edu">Yao.298@osu.edu</a></td>
<td>(614) 292-5044</td>
<td>Data Collection</td>
</tr>
</tbody>
</table>