The Wisdom of Crowds
Collaboratively Developing and Establishing the Validity & Reliability of a Student Teaching Evaluation Form

Erica Brownstein, Carolyn Kaplan, and Xiangquan (James) Yao, (The Ohio State University)
Kristall J. Day (Ohio Dominican University)
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
Pre-Assessment

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

• Has the instrument been analyzed for **Validity and Reliability**?

• Do you have **comparison** scores (with other EPPs)?
Discuss with your neighbor

• What are five (5) 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

Wright State University
  T. Kahrig

Kent State University
  J. Arhar
  S. Turner

Bowling Green State University
  D. Gallagher

Wittenberg University
  S. Brannan
  T. Whitlock
How was the CPAST Form developed?

**Year 1**
- Initial form developed and implemented at one EPP

**Year 2**
- Faculty/staff from 8 EPPs revised form
- Piloted with candidates from 8 EPPs and 322 candidates
- Validity & reliability analyses conducted

**Year 3**
- Faculty/staff from 10 EPPs revised form
- Implemented with 23 EPPs and 1203 candidates
- Validity & Reliability analyses conducted

*Based on (Ball & Forzani, 2010; Council of Chief State School Officers, 2011; Danielson, 2011; Gargani & Strong, 2014; Marzano, Pickering, & Pollock, 2001; SCALE, n.d.a, among others)
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

- Collaboration with 26 institutions
Resources

• Required CPAST Form Training for Supervisors
  • Required 80% score on supervisor training quizzes
  • Initial training ~90 minutes (self-paced)
  • Refresher training ~30 minutes (self-paced)

• Optional Cooperating Teacher/Student Teacher Training
  • ~20 minutes (self-paced)

• “Look Fors” Document
  • Elaborates on qualities and behaviors for a given level of performance, describing
    • Sources of Evidence
    • Possible Evidence
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

Data plus
- three observations
- IRR Supervisor final evaluation

IRR IHEs
# 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>

*Note - Sample includes representation from a variety of licensure areas, placement types, race, grade level, and gender.

Detailed demographic data available upon request.
Types of Validity and Reliability Analysis Conducted

Validity
- Content Validity
- Construct Validity
- Concurrent Validity

Reliability
- Test-retest Reliability
- Internal Consistency Reliability
- Interrater Reliability

- Addressed areas in the CAEP Evaluation Framework for Assessments
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).
Concurrent Validity

- Concurrent Validity examined by correlating CPAST scores with
  - state standardized professional knowledge and skills test scores and
  - edTPA scores

- No significant correlation between
  - CPAST Form and state professional knowledge and skills test
  - CPAST Form Disposition Rows and edTPA

- Significant correlation between
  - CPAST Form Pedagogy Rows and edTPA
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 *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*
Test-Retest

- Test-retest reliability examined by correlating Pedagogy and Disposition scores in the
  - midterm and
  - final term under the CFA model.

- Pedagogy & Disposition *significantly correlated* at midterm and final term
  - (Pedagogy =.752; Disposition =.845)
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.
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>

Also, see us at Spring 2017 CAEPCon March 24 at 11 am
And Fall 2017 CAEPCon September TBD.