Operationalizing VALUE Rubric dimensions
Case studies on calibration activities to enhance inter-rater agreement

**Introduction and Rationale**

- A trend is emerging in which, **STs** have been augmented with what some call a more authentic assessment method of rating **performance-based** student work (Rhodes, 2010).
- Unlike **ST**, **PBA** often takes the form of actual samples of student work (e.g., written student essays, written evaluative summaries of lab experiments, and oral presentations).
- While **PBA** has the advantage of authenticity, they also have an inherent disadvantage related to the difficulty of rating the student work samples.
  - Norming the test scores to specific student populations could suggest a pathway to comparisons based on equity and excellence (Chun, 2002).
- In contrast, **PBA** often requires the introduction of a **rubric**
  - to add consistency to the assessment of levels of performance
  - to assign points for comparisons.
- Montgomery (2002) recommended the use of **rubrics** for assessing authentic student work because it is a tool that communicates expectations about the quality or levels of applying of what was learned in order to implement the completion of the assignment.
- Yet, further research is needed to establish an **evidence-based approach for reliable and valid ratings by faculty**.

- Government policies ensure the quality of education in the USA
  - NCLB – states must measure student progress and a reliance on **standardized tests (ST)** grew.
  - ESSA – options for multiple measures, including merging **ST** and **performance based assessment (PBA)**.
  - **Across the country**, educational objectives broadly include critical thinking, communication, empirical and quantitative skills, teamwork, social responsibility, and personal responsibility.
  - **In our state**, the Texas Higher Education Coordinating Board (THECB) requires that all two-year and four-year educational institutions submit assessment reports for six objectives within general education courses approved and designated as the Texas Core Curriculum (THECB, 2011).
**Introduction and Rationale**

Examples: evidence-based solutions, implementation science, 17-yr avg

- A multi-state collaborative involving more than 60 institutions in 9 states has begun to report their findings (Lederman, 2015, Peden, 2017).
- Establishment of IU VALUE Institute
- Published studies have begun to report the construct validity of the VALUE rubrics (Gray, Brown, & Connolly, 2017).

**Learning Objectives**

After completing this session, participants will be able to

- name two best practices for a PBA scoring day as illustrated from case studies
- name two qualities to look for in a faculty facilitator and/or PBA raters applying a rubric to written student work
- recognize a rubric dimension that would benefit from operationalization using common language

**Introduction and Rationale**

This all points to the need for a registry of evidence-based approaches to ensure that the PBA rating data has significant agreement between raters prior to analysis.

- We are testing an approach of rater calibration that increases consensus and operationalizes the language in the rubrics that measure each dimension.
- We are currently testing and adapting this approach in order to highlight significant improvement in inter-rater agreement.
- This presentation will examine findings from Case Studies involving two signature assignment rating days and the utility of calibration.

**Learning from Case Studies**

**Case Study 1**
Using the Intercultural Knowledge and Competence VALUE Rubric to rate one Texas Core Curriculum objective: Social Responsibility

**Case Study 2**
Comparing the calibration process using two approaches and two teams of raters scoring the same set of student papers ($n = 76$)
Case Study 1 - Best Practices (logistics)

- Facilitator – leads faculty calibration
- 20 qualified faculty raters, by invitation
- Tables configured in a horseshoe
- One table centered inside the horseshoe
- Copies of the VALUE rubric
- Copies of two “anchor papers”
- Group reads the “anchor paper” silently
- Discussion led by facilitator
- SPSS ready laptop

Facilitator – leads faculty calibration
8-5 with working breakfast & lunch
assistant keeps papers moving
Tables configured in a horseshoe
One table centered inside the horseshoe
Copies of the VALUE rubric
Copies of two “anchor papers”
Group reads the “anchor paper” silently
Discussion led by facilitator
SPSS ready laptop

Demonstration

Instructions -

- Read the de-identified sample paragraph from authentic student work
- Read the rubric for the dimension we are rating
- Rate the dimension
  - Show of hands “how many awarded this a 4?”
  - Determine the value awarded by the major and ask for the reason(s) that it was a 1, 2, 3, or 4?
  - Raters make notes on the rubric for any operational language agreed upon by the group.

Demonstration 1

“...she expected to come to America. Something she received that she was denied in her home country was a sense of safety. Women, she said are targeted more often, regardless of race. Here, she isn’t afraid to walk alone, or to live away from her parents. This made me think about the increasing awareness on sexual assault and how much worse it must be in a third world country.

In her view, something she did not get was that each community has its unique culture. Here, she is able to walk alone, or to live away from her parents. This made me think about the increasing awareness on sexual assault and how much worse it must be in a third world country.”

Intercultural Knowledge and Competence

<table>
<thead>
<tr>
<th>VALUE RUBRIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competence</td>
</tr>
<tr>
<td>--------------</td>
</tr>
</tbody>
</table>
| Knowledge | Cultural/affirmation | - Articulates insights into own cultural roots and bases (e.g., dancing, weaving, cooking), shows evidence of broader cultural perspectives (e.g., in a college setting)
| - Recognizes new perspectives about one’s own cultural roots and bases (e.g., in a different college setting)
| - Identifies own cultural roots and bases (e.g., in a diverse college setting) |
| | 1 | 3 | 5 |

2017 SACSCOC Annual Meeting
Dallas, Texas
What we learned

Guided discussion of Best Practices

• Common sense approach that supports analyses of ratings if inter-rater agreement assumptions are met
• Facilitator qualities: skilled consensus builder, comfortable with rubrics, listener, respected faculty member, not necessarily the professor whose papers are being rated (combats notion of grading, promotes rating using the rubric)
• Room setup matters
• Let the calibration process last as long as necessary
• Encourage raters to circle evidence on "Anchor Papers" during calibration
• Papers that don’t align with the rubric frustrate raters
• Applying a rubric to assignments across disciplines = professional development

Case Study 2
Comparing different calibration processes (Groups 1 & 2) using the same rubric and same papers

Group 1 – same process approach described in Case Study 1

Group 2 – different process approach
• Led by 2 high-level administrators (one a former professor)
• 50+ raters, invited to an out-of-town training day on weekend (noon - 5pm with working lunch)
• More than a dozen circular tables with 8 raters seated at each
• Facilitator table & podium centered along the long wall in the rectangular room, projection screen on the narrow wall
• Copies of rubric and >five "anchor papers" in center of each table
• Facilitator-led discussion as before, during ~4-hour session
• Faculty rated papers at a later date (Range = 1 to six weeks later) via an on-line content management system

Two qualities required of raters

• Ability to think outside the box

• Willingness to follow evidence-based practices in applying a rubric to authentic student work

Demonstration

Two activities you can do with your raters that illustrate and highlight these required qualities
Connect all dots with four straight lines without lifting your pencil from the paper.

Think outside the box!

Art and Science of Rubric Use

"Not everyone is comfortable with qualitative, some people come to rating day asking what is a rubric?"
"Encourage your raters to circle evidence of the outcome on the rubric – why they gave the essay the rating they did"
"Spend enough time calibrating – tendency to want to begin – get nervous that you won’t finish all the essays... need to get ‘er done"
Debrief after rating is concluded – “what have you observed about the achievement of student learning for Social Responsibility?”

Rubrics help students
1. Understand assignment expectations and components
2. Improve their work through detailed feedback
3. Become more aware of their learning process... and progress

Rubrics help instructors
1. Assess assignments consistently
2. Give timely, effective feedback and promote SLO
3. Clarify assignment expectations and components
4. Refine teaching skills by evaluating rubric results

2017 SACSCOC Annual Meeting
Dallas, Texas
Key take-away points

“Cross disciplinary differences are overcome by effective calibration”

“Raters need a willingness to apply a rubric to disciplines not their own

• the ability to think outside the box
• can follow a protocol and reach a consensus with others”

The interval between calibration and rating matters!

References