**Introduction**

- This study is part of the core research agenda of the five-year, Goal 3 grant funded by the IES (# R305A100670) “Improving the Teaching and Learning of English Language Learners: The Instructional Conversation Model”
- As part of the assessment of fidelity of implementation (FOI) our research team developed a 20 items rubric to video-analyze teaching practices among treatment and control teachers.
- Bias in evaluation or rating in video review of teaching is evident in many studies.
- Some of it is inevitable given the use of video review for professional development purposes and the reflexive nature of that rating process.
- Other sources of bias are driven by competition for resources or reputation.
- Attempts to mediate this bias further via instrument validation and revision have been poorly documented.
- We address the latter issue using an Instrumental Iteration Toward Fidelity Of Implementation
- We aim at showing ways to improve the evaluation rubric while relying on raters with no vested interest in the individual outcomes or pressure to inflate scores.

<table>
<thead>
<tr>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>We rely on an iterative process to refine an existing 20-item evaluation rubric for rating pedagogical intervention in video-recorded 3rd and 5th grade lessons across subject types.</td>
</tr>
<tr>
<td>The goal is to increase reliability while potentially decreasing the number of items that may be used in future research and in the second and third waves of implementation of our current grant.</td>
</tr>
</tbody>
</table>

**Data and Analytic Strategy**

- **Participants**
  - The participants in this study are the video raters and teachers in treated and control settings.
  - The original video rating instrument was a 20 item checklist of yes/no binary outcomes.
  - A total of 25 videos were rated by 7 raters.
  - Raters were instructed to check an item should compelling evidence of its existence be shown in the rated video, regardless of teacher treatment status
- **Tetrachoric correlation**
  - Due to the binary nature of each item, we relied on tetrachoric correlation (Pearson, 1900) designed for binary data
  - The tetrachoric correlation coefficient is an excellent measure of rater agreement (Bonnet & Price, 2005)
  - The dimension reduction was implemented with factor analysis that used the matrix of tetrachoric correlation coefficients as input and Varimax rotation (Lieberman, 1970)
- **Cohen’s Kappa**
  - In addition to actual ratings per each rater, we have an aggregate measure that adds their total number of marked items, thus constructing a scale that ranges from 0 to 20.
  - Cohen’s Kappa compares observed agreement between raters with the level of agreement that is expected to occur randomly by analyzing the following two quantities
  - Assuming that raters consistently assessed teachers’ performance, we would expect there to be a high level of reliability between any given numbers of raters’ scores that assessed a teacher, regardless of treatment status.

**Findings / Results:**

- The tetrachoric correlation found is 0.5086 (p. < .001) expressing overall rater association.
- The importance of this analysis, however, relies on observing which items are highly correlated.
- For example, we observed that raters’ identification of teacher encouraging student conversation is highly correlated with raters’ identification of actual student dialogue (rho= .80).
- However, the former is uncorrelated (rho= .003) with raters’ identification of teacher listening to assess student levels of understanding.
- Kappa coefficient of four raters analyzing seven videos was 0.586 (p ¡ 0.001).
- Raters responses were correlated above chance regarding the evaluation of the IC

**Discussion and Implications**

- Promotion of dialogue does not necessarily impact teachers’ implementation of the listening component of the IC.
- Yet, the identification of listening is correlated with (rho= .43) the identification of teachers’ connecting lesson content to student experiences, which is an important component of the IC model.
- Take home point?
- The three techniques employed show evidence of FOI
- The evaluation of the IC implementation can be reliably assessed with the current version of the scale
- However, modification to the instrument based on the findings from the tetrachoric correlation and factor analyses may improve interrater reliability and provide a better assessment of FOI.