Exploring the Reliability of Generic and Content-Specific Instructional Aspects in Physical Education Lessons: Insights from an Exploratory Study

Charalambos Y. Charalambous, Ermis Kyriakides, Niki Tsangaridou, Leonidas Kyriakides
Structure of Presentation

- The importance of classroom observations in measuring teaching quality
- Generalizability theory and its contribution to exploring issues of reliability
- Research Questions
- Methods
  - Instrumentation (Generic and Content-Specific Instructional Aspects)
  - Participants, Rater Training
  - Data collection and data analysis
- *Selected* findings
- Discussion and Implications
Different approaches for measuring teaching quality:

- Teacher ratings (Kunter & Baumert, 2006)
- Student ratings (Fauth et al., 2014)
- Teacher logs (Rowan, Harrison, & Hayes, 2004)
- Instructional artifacts (Martínez, Borko, & Stecher, 2012)
- Classroom observations (Wragg, 2012)

The potential of classroom observation

- Observations yield more reliable measures, as they can avoid many of the biases of self-report data (Strong, 2011)
Comparison of G-theory and CTT:

\[ X = T + E \]

Observed score \quad True score \quad Random error

CTT:
- Random and Systematic sources of error

G-theory:
- Random sources of error
  - F1
  - F2
  - F3

Systematic sources of error

\[ F1 \leftrightarrow F2 \]
The G-theory framework

- **D-Studies:**
  - D-studies: thought experiments that help design future studies to maximize reliability in cost effective ways

- **Factors influencing classroom observation estimates:**
  - Observational instrument itself
  - Recruitment and training of raters
  - The scoring design (e.g., the number and the length of observations, the number of raters, the sequence of observations)
  - ... (Casabianca et al., 2013; Hill, Charalambous, & Kraft, 2012; Kane & Staiger, 2012)
Significance of present study

- No studies have so far utilized this framework to examine the reliability of estimates of teaching quality obtained from classroom observations of PE lessons.

- PE differs significantly from other content-areas
  - PE often focuses on different learning outcomes (psychomotor instead of cognitive).
  - Lessons are conducted in open-space within which students are constantly moving; hence learning might be affected by weather conditions or the possibility of an injury (Lindsay, 2014).

- Generic vs Content-Specific instructional dimension
Generic and Content-Specific Instructional Aspects

- **Generic Instructional Aspects**
  - Instructional features that cut across different disciplines
  - They are important for teaching, regardless of the subject matter that gets taught (e.g., time and classroom management)

- **Content-Specific Instructional Aspects**
  - Instructional features that are particularly relevant to specific content-areas
  - e.g., the use of demonstration for the desired movement skills for the discipline of PE
Research Questions

- What is the optimal combination of lesson observations and raters coding these lessons needed to yield reliable estimates of teachers’ practice in PE?

- Does this optimal combination differ across generic and content-specific aspects of instruction?
**Methods: Instrumentation**

- **Sampling instrument: High Inference Rubric**

  - 19 statements

  - **F1: Classroom Management** (4 statements)
  - **F2: Time Management** (3 statements)
  - **F3: Task Progression** (i.e., informing, refining, extending, applying) (4 statements)
  - **F4: Demonstration of Desired Movement Skills** (5 statements)
  - **F5: Quality of Student Practice** (3 statements)
  - **F6: Generic Instructional Aspects**
  - **F7: Content-Specific Instructional Aspects**

  **CFA Analysis**
Methods: Participants, Rater Training and Data Collection

- **Participants**
  - 49 generalist teachers who taught PE to 3rd to 5th elementary school students

- **Raters and Rater Training**
  - Four second-year master’s students in PE
  - Observing and coding videotaped and actual PE lessons
  - Certification when at least 80% agreement was obtained with master-coder ratings

- **Data Collection**
  - Three scheduled observations of 40-minute typical daily lessons of PE for each teacher
Methods: Design and Data Analyses

- **Design**
  - Two-facet design: Persons x Raters x Occasions
  - Analyses of the seven factors (five first-order and two second-order)

- **Data Analyses**
  - GENOVA software
  - G-Study: Partitioning the variance into three components under consideration (i.e., Person, Rater, and Occasion) and their interactions
  - D-Studies: Altering the number of raters and the number of occasions for each factor to achieve at least 65% reliability
**Selected Findings (1)**

**Variance Decomposition for the Seven Factors of the High-Inference Instrument**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>F1*</th>
<th>F2*</th>
<th>F3*</th>
<th>F4*</th>
<th>F5*</th>
<th>F6*</th>
<th>F7*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers (t)</td>
<td>38.94</td>
<td>38.69</td>
<td>35.52</td>
<td>44.61</td>
<td>14.37</td>
<td>41.98</td>
<td>47.29</td>
</tr>
<tr>
<td>Raters (r)</td>
<td>0.00</td>
<td>0.00</td>
<td>1.27</td>
<td>1.16</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Occasions (o)</td>
<td>0.00</td>
<td>2.16</td>
<td>0.00</td>
<td>0.25</td>
<td>0.00</td>
<td>2.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Teachers × Raters (t × r)</td>
<td>2.55</td>
<td>5.24</td>
<td>1.21</td>
<td>1.86</td>
<td>4.21</td>
<td>4.82</td>
<td>1.66</td>
</tr>
<tr>
<td>Teachers × Occasions (t × o)</td>
<td>33.33</td>
<td>35.23</td>
<td>49.10</td>
<td>42.50</td>
<td>37.29</td>
<td>32.21</td>
<td>42.57</td>
</tr>
<tr>
<td>Raters × Occasions (t × o)</td>
<td>0.77</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.87</td>
<td>0.36</td>
<td>0.05</td>
</tr>
<tr>
<td>Teachers × Occasions × Raters (t × o × r), residual</td>
<td>24.42</td>
<td>18.68</td>
<td>12.89</td>
<td>9.58</td>
<td>43.25</td>
<td>18.61</td>
<td>8.43</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

*Average of items of each factor*
Selected Findings (2)

Factor 6: Generic Instructional Aspects

Factor 7: Content-Specific Instructional Aspects
Selected Findings (3)

Factor 3: Task Progression

Factor 5: Quality of Student Practice
Importance of exploring reliabilities yielded from observational rubrics using the G-theory framework

Reliabilities cannot and should not be taken for granted: they are the composite of different components within an observational system

Different dimensions might exhibit different reliabilities

Implications

- Rater training and certification
- Appropriateness of existing teacher evaluation approaches?
Questions?

Comments?

Suggestions?
Thank you for your attention!

Contact information:

- Charalambos Y. Charalambous  
  cycharal@ucy.ac.cy
- Ermis Kyriakides  
  kyriakides.ermis@ucy.ac.cy
- Niki Tsangaridou  
  edniki@ucy.ac.cy
- Leonidas Kyriakides  
  kyriakid@ucy.ac.cy
References


