Earli 16th Biennial Conference
Earli 2015

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

August 29, 2015

University of Cyprus
Department of Education

Session 19: School Effectiveness

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

The Role of Observations for Measuring Teaching Quality

- ***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)

Introduction: G-Theory and CTT

Comparison of G-theory and CTT:

$$X = T + E$$

Observed score

True score

Random error

CTT:

Random and Systematic sources of error

G- theory:

Random sources of error

+

Systematic sources of error

F1

F2

F1 ⇔ F2

F3

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

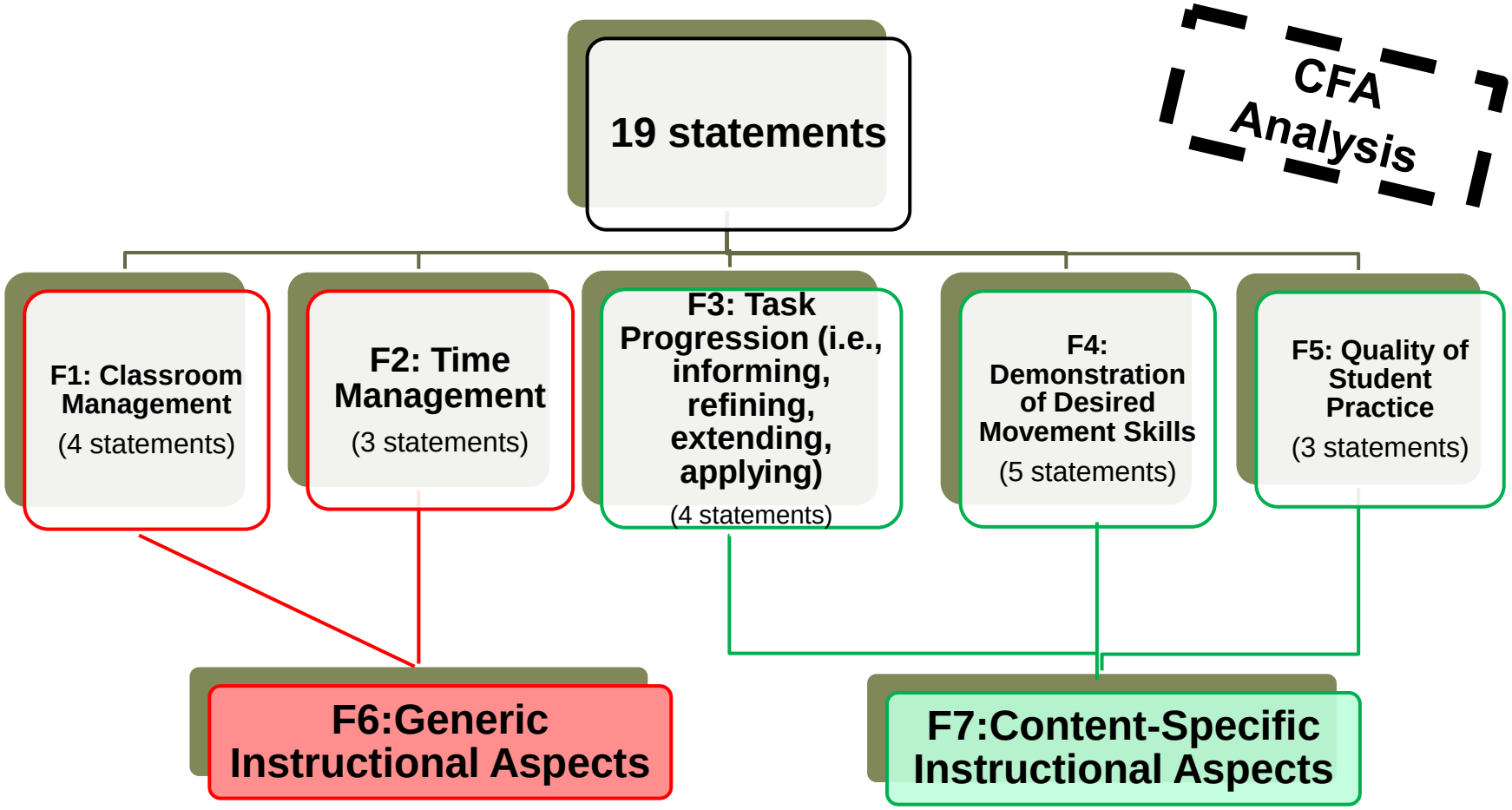
Research Questions

- 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***



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

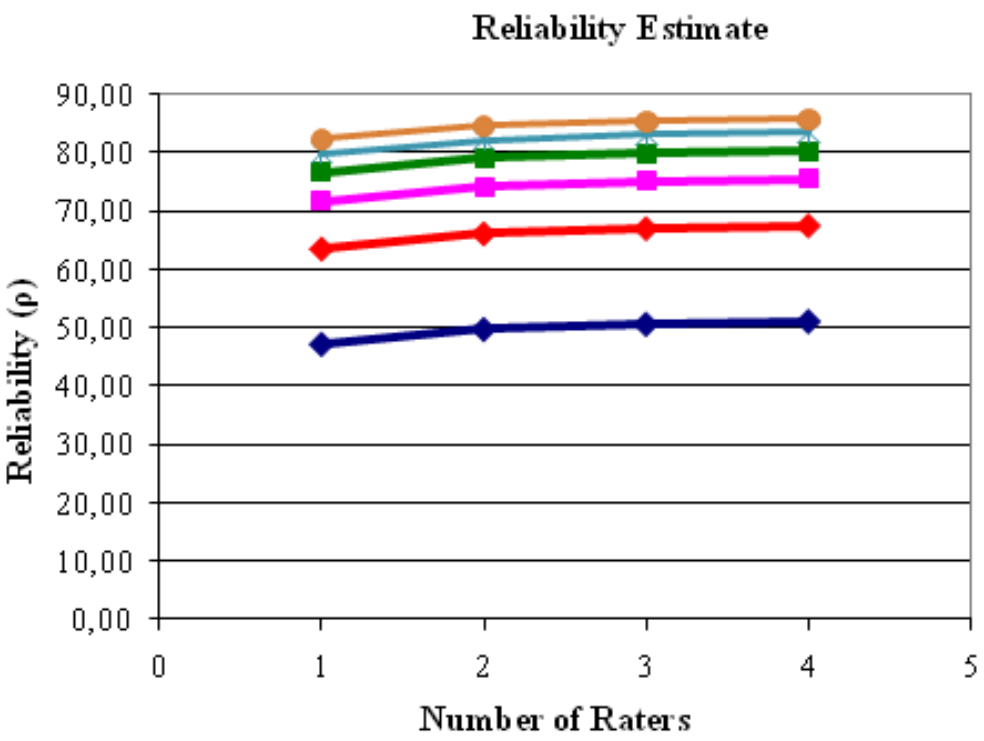
Selected Findings (1)

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

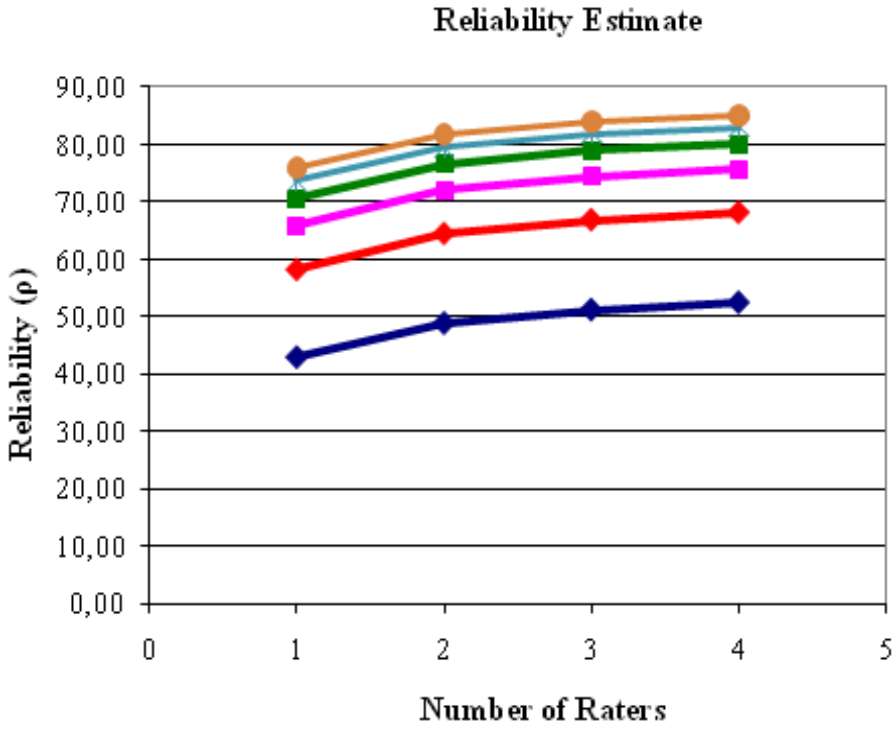
	Factors						
Source of Variation	**F1***	**F2***	**F3***	**F4***	**F5***	**F6***	**F7***
Teachers (t)	38.94	38.69	35.52	44.61	14.37	41.98	47.29
Raters (r)	0.00	0.00	1.27	1.16	0.00	0.00	0.00
Occasions (o)	0.00	2.16	0.00	0.25	0.00	2.02	0.00
Teachers × Raters (t × r)	2.55	5.24	1.21	1.86	4.21	4.82	1.66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

**Average of items of each factor*

Selected Findings (2)

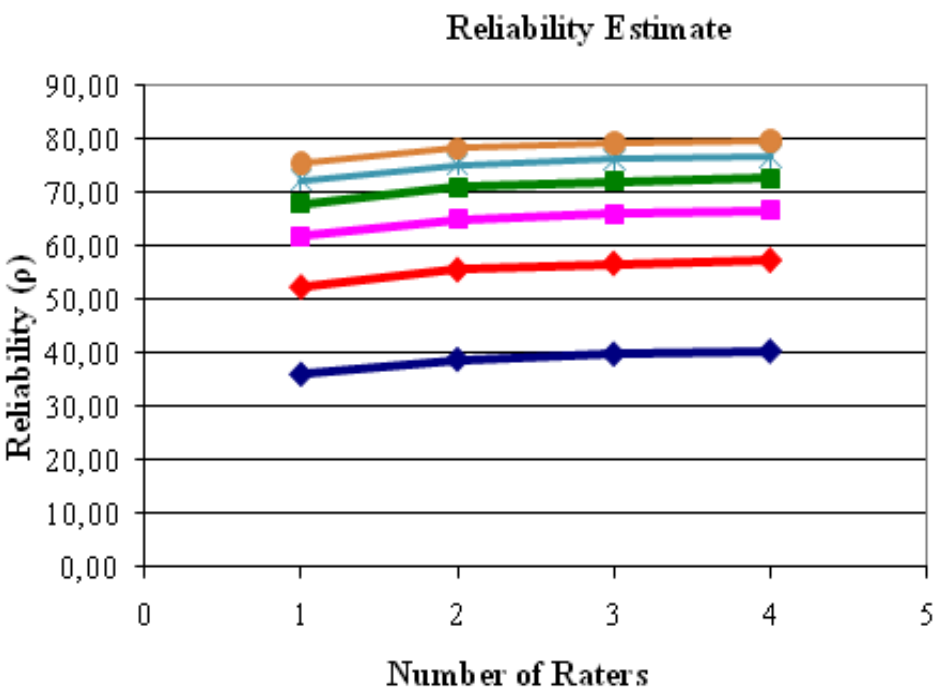


Factor 6:Generic Instructional Aspects

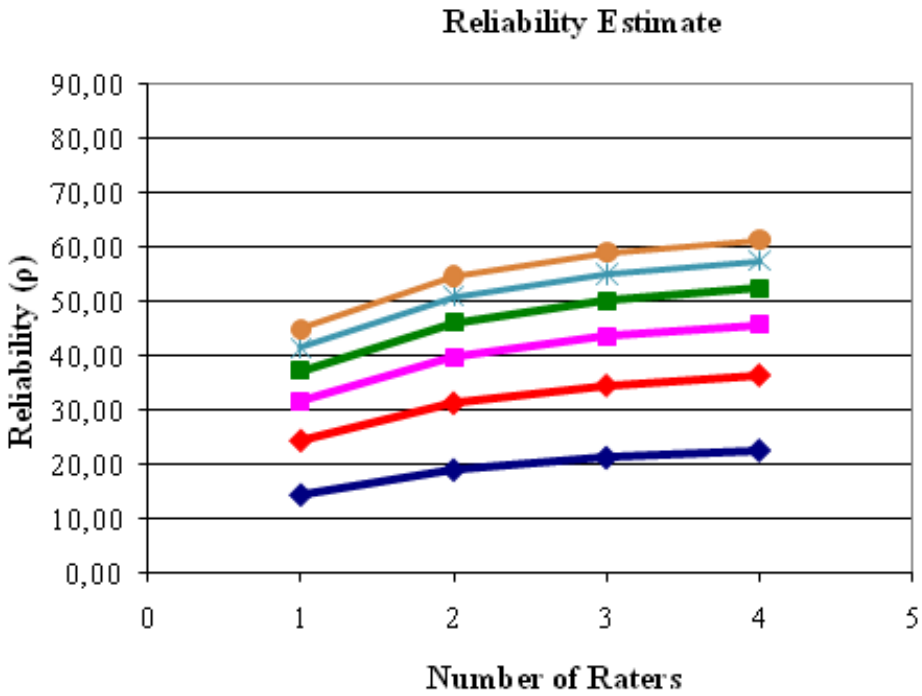


Factor 7:Content-Specific Instructional Aspects

Selected Findings (3)



Factor 3:Task Progression



Factor 5:Quality of Student Practice

Discussion

- 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

- Casabianca, J. M., McCaffrey, D. F., Gitomer, D. H., Bell, C. A., Hamre, B. K., & Pianta, R. C. (2013). Effect of observation mode on measures of secondary mathematics teaching. *Educational and Psychological Measurement*, *73*(5), 757-783, doi:10.1177/0013164413486987.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1-9, doi: 10.1016/j.learninstruc.2013.07.001.
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, *41*(2), 56-64, doi: 10.3102/0013189X12437203.
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle: Bill & Melinda Gates Foundation. Retrieved November 30, 2012, from http://www.metproject.org/reports.php
- Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environment Research, 9,* 231-251, doi: 10.1007/s10984-006-9015-7.
- Lindsay, E. L. (2014). Effective teaching in physical education: The view from a variety of trenches. *Research Quarterly for Exercise and Sport, 85*(1), 31-37, doi: 10.1080/02701367.2014.873330.
- Martinez, J. F., Borko, H., Stecher, B. M. (2012). Measuring instructional practice in science using classroom artifacts: Lessons learned from two validation studies. *Journal of Research in Science Teaching, 49*(1), 38-67, doi: 10.1002/tea.20447.
- Rowan, B., Harrison, D. M., & Hayes, A. (2004). Using instructional logs to study mathematics curriculum and teaching in the early grades. *The Elementary School Journal*, *105*(1), 103-127.
- Strong, M. (2011). *The highly qualified teacher: What is teacher quality and how do we measure it?* New York, NY: Teachers College Press.