

Using Classroom Observation Research to Inform Debates about Teaching Effectiveness

Perspectives

Urgent debates about teaching effectiveness pervade media channels (Gladwell, 2008; Meyers, 2010), policy documents (Gordon, Kane, & Staiger, 2006), and educational research (Pianta & Hamre, 2009; MET, 2011), as we move into an era of high-stakes testing with teacher, student, and school-level accountability. Pianta and Hamre (2009) argue for the importance of classroom observation research to the field of education, while noting that instruments appropriate for high school are “the exception rather than the norm” (p. 111).

Our team of scientists, educational researchers, and master teachers collaboratively developed a classroom observation protocol to evaluate graduates of a nationally-recognized secondary math and science teacher preparation program. This program couples a strong focus on content knowledge with intensive classroom and field training in research-based teaching practices (e.g., NRC, 1996; NCTM, 1991; Bransford, Brown, & Cocking, 2000). A discussion of the effort to develop a protocol with these foci is presented here.

Theoretical Framework

Shulman (1986; 1987) describes pedagogical content knowledge (PCK) as professional knowledge of the domain that supports teachers in promoting understanding. Shulman differentiates PCK from both content knowledge and content-general pedagogical approaches. Ball, Thames, and Phelps (2008) further subdivide content knowledge into common content knowledge (subject knowledge) and specialized content knowledge (subject knowledge related to practice, not necessarily known by experts), and differentiate these from PCK (knowledge relating to students and teaching).

However, despite decades of scholarship, there are no conclusive answers about what teaching practices are effective for promoting learning. Hiebert and Grouws (2007) reviewed the literature on

mathematics teaching, writing “documenting particular features of teaching that are consistently effective for students’ learning has proven to be one of the greatest research challenges in education” (p. 371).

Methods and Data Sources

We reviewed current mathematics-specific classroom observation instruments, and decided to modify a protocol developed by Horizons Research (1999). We refer to our instrument as the Mathematics Classroom Observation Protocol (MCOP). The MCOP contains 25 indicators rated on a 1-5 scale, with four sections: *Classroom Environment*, *Lesson Structure*, *Implementation*, and *Mathematics Content*. Detailed rubrics and examples were developed for each indicator.

The MCOP was used by 99 trained raters on 994 video 4-8 mathematics lessons taught by a national sample of 250 teachers from six school districts in six states. Rater training included video viewing and collaborative discussion. Descriptive analyses of indicator ratings, along with statistical analyses of inter-rater reliability, internal consistency, and factorial structure of the MCOP were conducted. Our primary goal was to develop a reliable and valid observation protocol which could be used in a variety of contexts and by different stakeholders. We also wanted to contribute to the national conversation about effective teaching in mathematics.

Results

Cluster Analysis

Factorial analyses of the MCOP revealed 4 clusters (Figure 1). The first cluster related to how the teacher promoted surface-level engagement. These indicators mainly described general pedagogical skills, such as classroom management. Many videos scored highly on these indicators (Figure 2, blue).

The second cluster related to teacher efforts to build conceptual understanding of significant mathematics concepts. These included behaviors like fostering meaningful student contributions and using higher-order questioning techniques, similar to Ball and Forzani’s (2009) “high leverage” teaching practices. As shown in Figure 2 (red), few teachers in this national sample consistently scored well on

these indicators, suggesting that teacher preparation programs may need to re-conceptualize how these skills are taught.

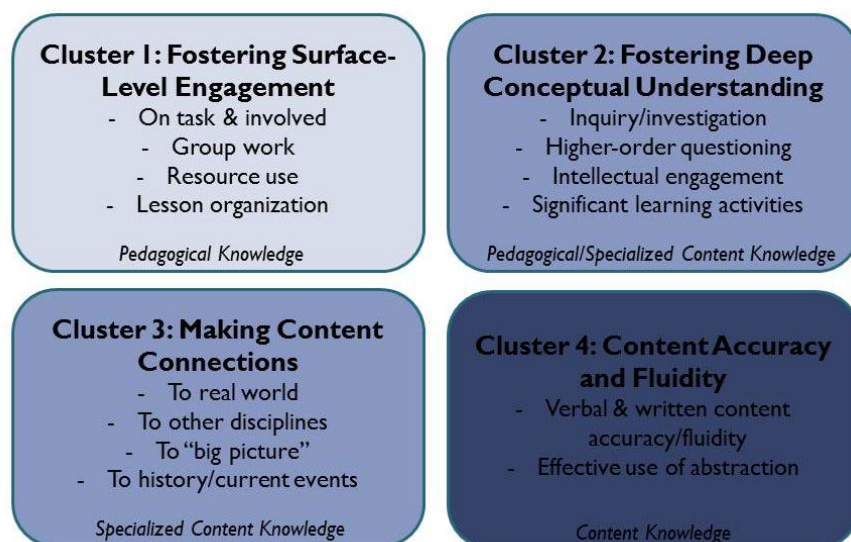


Figure 1. Four clusters revealed by factor analysis of MCOP indicators

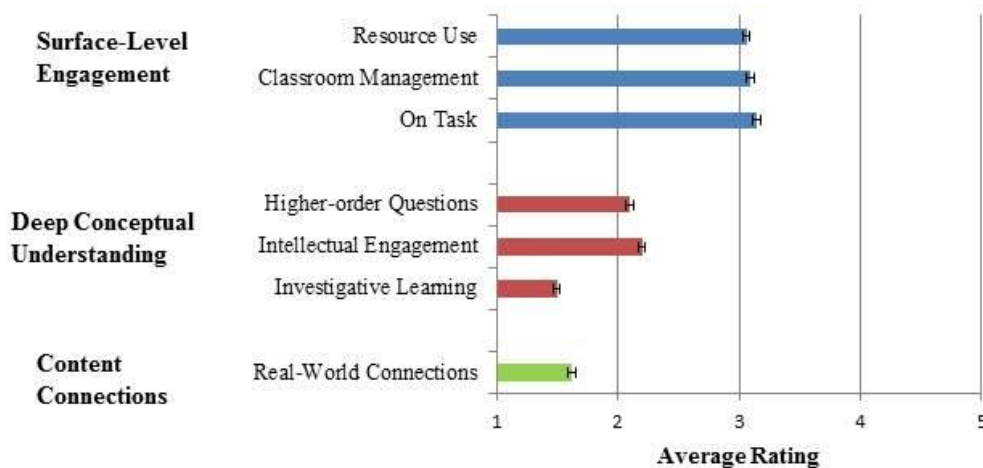


Figure 2. Mean scores ($N = 994$ videos) with SEM bars for selected indicators in three clusters

The third cluster, relating to specialized content knowledge, assessed how the teacher connected mathematics to the world, to history and current events, and to the “big picture” of the discipline. As

suggested by Figure 2 (green), these behaviors were rarely seen. Although such connections are highly emphasized in our preparation program, they may not be part of the normal training of teachers or mathematicians. These teaching behaviors may be critical to fostering intrinsic motivation and promoting integrated, applied understandings of the subject matter.

The fourth cluster focused on content knowledge, including communicating content accurately. One unexpected finding was the prevalence of content mistakes. In nearly half (48%) of the lessons, the teacher communicated content that was incorrect or highly problematic (Table 1). The prevalence of content mistakes challenges notions that content-general instruments or raters not highly trained in the content area can evaluate effective teaching.

Description of Content Issue	Examples (fictionalized to protect teacher privacy)
Inaccurate	“A marker has no volume because you can’t fit anything inside of it.” “You can use the commutative property of division.”
Grade Level Issues	$\sqrt{x^2} = x$ $x(x-1)/x = (x-1)$
Formulaic/ Not Generalizable	“Remember when solving proportion word problems, that ‘of’ means divide.” “For fractions, remember the denominator is always the bigger number.”
Problematic	- Using the same variable name in single problem to represent different quantities - Misuse of equals sign: $7 + 3 = 10 + 2 = 12$ Dan = 4

Table 1. Examples (fictionalized to protect anonymity) of incorrect content communicated by teachers

Rater Reliability

A key issue we confronted was determining who is qualified to evaluate teaching using the MCOP. The weighted kappa for each of the 99 raters was used as a dependent measure in a regression model. Results showed that mathematics teacher raters were significantly ($p < .05$) more reliable than science teacher raters. However, this was because science raters had higher standards for surface-level engagement. Raters with content or mathematics education undergraduate degrees were significantly ($p <$

.01) more reliable than raters with other degrees. This seemed to result from raters in the latter group having difficulty catching content mistakes.

Significance

The MCOP was designed to assess how teachers foster conceptual understanding and how content expertise contributes to effective teaching. Using a national sample of middle school math lessons, we show how standards-based teaching practices forwarded by reform movements are rarely seen, and that communication of inaccurate content is a regular occurrence. This suggests that the MCOP is a valuable tool to assess the impact of teacher preparation programs, and that the clusters underlying the indicators may be important to discussions about effective teaching.

The MCOP was originally divided into 4 sections (Classroom Environment, Lesson Structure, Implementation, and Mathematics Content); however analyses showed that such surface-level features of the lesson may not be particularly useful in conceptualizing effective instruction. We show the importance of rater expertise in classroom observation. Finally, we demonstrate that teachers can be trained to make critical distinctions about effective teaching when using the MCOP, suggesting such protocols may be valuable for professional development.

References

- Ball, D., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389-407.
- Ball, D., & Forzani, F. (2009). The work of teaching and the challenge of teacher education. *Journal of Teacher Education, 60*(5), 497-511.
- Bransford, J., Brown, A., & Cocking, R. (2000). *How People Learn*. Washington D.C., National Academy Press.
- Gladwell, M. (2008). Most likely to succeed: How can we hire teachers when we can't tell who's right for the job? *The New Yorker*. New York.
- Gordon, R., Kane, T., & Staiger, D. (2006). Identifying effective teachers using performance on the job. Discussion Paper 2006-01. The Hamilton Project.
- Hiebert, J., & Grouws, D. (2007). The effects of classroom mathematics teaching on students' learning. In D. Grouws (Ed.), *Handbook of Research on Mathematics: Teaching and Learning* (pp. 371-400). New York: MacMillian.
- Hill, H., Rowan, B., & Ball, D. (2005): Effects of teachers' mathematics knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371-406.
- Horizons Research Inc. (1999). Local Systemic Change through Teacher Enhancement Classroom Observation Protocol. Retrieved June 2009 from <http://www.horizon-research.com/instruments/lsc/cop.php>.
- Measures of Effective Teaching (MET) Project (2011). Accessed 14 June 2011 from <http://www.metproject.org/>
- Meyers, J. (2010). Evaluating teacher effectiveness is evolving. *The Dallas Morning News*. Accessed 14 June 2011 from <http://www.dallasnews.com/news/education/headlines/20101011-Evaluating-teacher-effectiveness-is-evolving-5189.ece>
- National Council for Teachers of Mathematics (1991). *Professional standards for teaching mathematics*. Reston, VA.

National Research Council, National Academy of Sciences. (1996). *National science education standards*. Washington, DC.

Pianta, R., & Hamre, B. (2009). Conceptualization, measurement, and improvement of classroom processes: Standard observation can leverage capacity. *Educational Researcher*, 38(2), 109-119.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *American Educational Research Association Journal*, 15, 4-14.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-27.