

Standardization in fluid contexts.

Methodological constraints in large-scale video studies of teaching quality.

Introduction

There is a growing consensus on the importance of teaching quality as a central factor for student learning (Darling-Hammond, 2017; Hattie, 2009). Accordingly, educational researchers aim to identify factors that correlate with more efficient learning, to understand how instructional factors interplay with different contextual factors, and to identify areas for professional development of teachers. These are also core ambitions in the newly formed Nordic research centre, Quality in Nordic Teaching (QUINT), funded by NordForsk.

In order to make cross-country comparisons and utilize powerful statistical analysis, large-scale datasets (from observations, surveys, student outcomes etc.) are necessary. Part from being both time consuming and expensive, large-scale data collection also requires a great deal of standardization in order to generate variables that are “clean” enough (i.e., free from contextual variation) to be subjected to statistical analysis. For instance, a model may suggest that observational data from one student group can be tied to average student gains over a school year in the same group.

In this presentation, we raise a set of methodological issues that relate to the standardization of large-scale observations of teaching. In addition, we discuss to what extent the fluidness of contexts such as class, lesson, subject, and material are characteristics of the Nordic school, and what it means for empirical research of teaching quality.

The LISA study in Sweden

The Linking Instruction and Student Achievement (LISA) study in Sweden replicates the research design of the Norwegian LISA study (Klette, Blikstad-Balas, & Roe, 2017), and aims i) to reveal how different features of instruction relate to student learning in Swedish and mathematics; ii) to identify differences and similarities between language arts and mathematics instruction in the Nordic countries; and iii) to explore the contextual validity of observational protocols such as the Protocol for Language Arts Teaching Observation (PLATO) (Grossman et al., 2013).

Classroom video observations and analysis by PLATO are combined with student survey of perceived teaching quality, and student gains over the course of one school year. PLATO scores from four consecutive lessons in each subject and group of students is used to represent the observed teaching quality. These data are analytically tied to average gains in each student group over one school year, and to the survey representation of perceived quality by the same student group.

Methodological constraints

Previous research has pointed to several methodological challenges in doing large-scale observations of teaching quality, such as theoretical fragmentation that impedes the integration of empirical results (Scheerens, 2016), reliability of coding teaching practices (Ho & Kane, 2013), and relevance of different student outcome measures (Grossman et al., 2014). These are all highly relevant factors and critical when analyzing teaching quality by inferential statistics.

In this presentation, we point to additional challenges that we have run across during data collection and preliminary analyses. They concern the fluidity of educational

settings that may interfere with the ambition of standardized observation and monitoring of students' learning progress. The following areas are discussed:

1. Flexible study groups over the school year. Two or three teachers collaborate in doing flexible grouping of students during different periods of the year, sometimes by ability.
2. Two-teacher systems including flexible classroom solutions. Two teachers take joint responsibility for a class. Division of instructional tasks and nature of interplay between teachers can vary. In order to mobilize resources, solutions such as large-group plenaries are sometimes needed to facilitate two teachers on other lessons.
3. Cross-disciplinary work (e.g., language arts + social science in combination). Teachers from different subjects periodically collaborate in thematic studies, thereby making borders between for subjects temporarily fluid.
4. Subdisciplinary work (e.g., rhetorics/communication cut out from the rest of language arts). Subdomains of a subject are cut out to be scheduled and taught separately.
5. Digitally distributed teaching (e.g., films, slide presentations, demonstrations), student work (texts, assignment responses), and examinations (tests etc.). Learning Management Systems are used to organise teaching, thereby making both instruction and student production less accessible for classroom observation.

Discussion

Classrooms are dynamic sites for teaching and learning, and traditional notions of stable correspondence between one teacher, one group, and one subject are not self-evident. The identified cases of mis-match between research design and empirical reality can to a large extent be understood in relation to ideas of professional collaboration as means for developing teaching quality. The observed practices are grounded in ideas about professional communities of teachers and widespread in Nordic schools. Yet, they will represent forms of contextual constraint when researchers try to capture data by protocols that expect one teacher-one group-one subject relationships.

These observations of mis-match call for further discussion:

- Standardization is necessary to accomplish comparability, but contextual differences might affect comparability when transferring research designs between contexts.
- How could this be managed in large-scale international video studies?
- What bias does it mean to exclude teachers and classrooms that does not fit to the presupposed conditions of one teacher-one class-one subject?
- How does the fluidity of the dynamic classroom affect the result?

References

- Darling-Hammond, L. (2017). *Empowered educators: How high-performing systems shape teaching quality around the world*. San Francisco: Jossey-Bass.
- Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value-added on multiple types of assessment. *Educational Researcher*, 43(6), 293-303.
- Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education*. 119(3), 445–470.

- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Ho, A. D., & Kane, T. J. (2013). The reliability of classroom observations by school personnel. Seattle, WA: Bill & Melinda Gates Foundation.
- Klette, K., Blikstad-Balas, M., & Roe, A. (2017). Linking Instruction and Student Achievement. A research design for a new generation of classroom studies. *Acta Didactica Norge*, 11(3), 1–19.
- Scheerens, J. (2016). *Educational effectiveness and ineffectiveness. A critical review of the knowledge base*. Springer Netherlands.