Paper Title: Studying Beginning Mathematics Teaching Using MQI, CLASS, and FFT

Author(s): L. Joy Johnson, Delena M. Harrison & Kolby J. Gadd

Session Title: Studying Beginning Mathematics Teaching Using MQI, CLASS, and FFT

Session Type: Brief Research Report

Presentation Date: Wednesday, April 13, 2016

Presentation Location: San Francisco, California

Studying Beginning Mathematics Teaching Using MQI, CLASS, and FFT

L. Joy Johnson, Delena M. Harrison & Kolby J. Gadd

University of Michigan

National Council of Teachers of Mathematics 2016 Research Conference

Abstract

Studies of student achievement suggest that early-career mathematics teachers are less effective than their more experienced peers. These findings and other concerns lead us to investigate the features of early-career mathematics teaching or what we call beginning mathematics teaching. Using hierarchical measurement and logistics regression models, we analyze teacher observation data from approximately 500 teachers in five districts to learn more about beginning mathematics teaching. Based on our experience with the Measuring Effective Teaching Longitudinal Database, we consider the usefulness of this database for secondary analysis and the optimum feature of a dataset for the secondary analysis of teaching.

Studying Beginning Mathematics Teaching Using MQI, CLASS, and FFT

Much of what is known about early career mathematics teaching is nestled among our knowledge of pre-service and in-service teachers and teaching (Richardson, 2001). Our colleagues Mann and Ball (2015) argue that studying the early years of mathematics teaching or what they call beginning mathematics teaching has its own merit. They note a trend of high turnover of teachers leading to a high proportion of early career teachers working in U.S. classrooms over the last two decades. Additionally, these early career teachers are more likely to be assigned to classes of higher proportions of high-poverty students and students of color, populations traditionally underserved by the U.S. public school system (Ingersoll, & Merrill, 2010) or referenced in discussions of achievement gaps (Vanneman, Hamilton, Baldwin Anderson & Rahman, 2009). Furthermore, Mann and Ball argue that the assignment of beginning teachers to low performing schools is particularly concerning because students of beginning teachers, on average, have lower achievement than students of more experienced teachers (Gordon, Kane, & Staiger, 2006).

We seek to learn more about beginning mathematics teaching in order to improve instruction and address achievement gaps in mathematical knowledge and inequalities in opportunities to learn using data collected by the Measures of Effective Teaching (MET) project. The research discussed in this paper considers the usefulness of the MET Longitudinal Database (MET LDB) for secondary analysis of mathematics teaching and seeks to better understand differences between beginning and experienced teaching through the measures of teaching practice that it contains. This brief paper presents a summary of our argument and preliminary findings.

Over the past four years, our research group, the Study of Beginning Mathematics Teaching[1] (SBMT) at the University of Michigan has been studying early career mathematics teaching in several ways. A primary focus of our work has been the study of video-records of mathematics instruction by teachers in their first three years of teachings. The qualitative analyses of this data has focused on coordination and working with a representation in mathematics instruction (Farmer, Mann, Hickman, & Snider, 2015). Simultaneous to this work, our team has considered what might be learned by studying the teaching of mathematics by beginners at scale. This second line of research, therefore, seeks to understand systemic differences between beginning and experienced mathematics teaching through studying records of practice. In particular, this line of inquiry sought to exploit the availability of the largest collection of data on teaching collected to date—the MET LDB. Thus, our work took on a dual purpose 1) to investigate differences between beginning and experienced mathematics teaching and 2) to consider the optimum features of a dataset of teaching practices conducive to large-scale, secondary investigations.

Our initial research using the MET LDB probed whether there are differences in the overall Mathematical Quality of Instruction (MQI) between beginning and more experienced teachers during mathematics instruction (Harrison, Gadd, Johnson, & Mann, 2015). We found that there is a minimum sample of beginning teachers of mathematics, but given the other exceptional features of this data, we concluded that it was a reasonable sample for an initial foray into studying beginning mathematics teaching. Furthermore, inside these sampling and

[1] The SBMT research group was supported by the National Science Foundation under grant DRL-1008317. We thank members of our research group for their help in thinking about these ideas: Deborah Ball, Mark Hoover, Hyman Bass, Lindsey Mann, Lauren Hickman, Esther Enright, Susanna Farmer, Sarah McNair-Wilson, Austin Raymond, and Rachel Snider. The opinions reported here are those of the authors and do not necessarily reflect the views of the National Science Foundation or our colleagues.

measurement issues, we found some evidence that beginning teachers were more likely to have a significantly lower MQI, a tool for measuring quality of mathematics instruction, than more experienced teachers. We found this even when controlling for behavior management and mathematical knowledge for teaching, two characteristics that might seem to distinguish beginning and experienced teaching. Because this research only considered underlying patterns across several elements of MQI, we expect there is more to learn about beginning teaching including the mechanism or instructional traits that contribute to these differences. The work reported here builds on this prior study and probes differences in the scores of beginning and experienced teachers for three teaching observation protocols, MQI, Classroom Assessment Scoring System (CLASS), and Framework for Teaching (FFT), during records of mathematics instruction and reports on how well this set of measures and the overall methods and study design used in the MET LDB work for studying mathematics teaching.[2]

Conceptual Framework

Our work is framed by the theoretical perspective that instruction consists of interactions between teachers, students, and content within an environment (Cohen, Raudenbush, & Ball, 2003; , 2001). The central idea is that teachers neither deposit information into students nor have ultimate control over the instruction that takes place. Instead, instruction is co-created as teachers and students work together on content, and learning occurs through these interactions. We apply this theory in our study of beginning mathematics teaching and hypothesize that instruction has unique patterns that are identifiable by teachers' position as beginners.

This theory of instruction is fundamentally useful for how we think about differences

---

[2] The current study builds on our prior study of beginning teaching using the MET LDB (presented at the 2015 National Council of Teachers of Mathematics Research Pre-session in Boston). We thank Lindsey Mann for her help in formulating our research questions and designs.

between mathematics instruction of beginning and experienced teachers. If we replace "teacher" in the instructional triangle with "beginning teacher" and in a separate triangle "teacher" is now "experience teacher," we hypothesize that each of these two groups of teachers interact with content, students, and the environment differently, particularly for instruction on mathematics content. Their approach, disposition, and/or knowledge for this work is likely different, but those differences are not well understood. We argue that given the increased concentration of beginning teaching in the current and future workforces and the opportunity gaps for poor and minority students, understanding these differences is of contemporary importance (Lankford, et al., 2002). Furthermore, a better understanding of what characterizes beginning mathematics teaching could inform both initial professional training and resources or structures designed to support beginning mathematics teaching once in service.

The MET LDB and Secondary Analysis: A Dual Problem

The MET Project collected the MET LDB seeking a better understanding of the utility of teaching evaluation methods as a means of feedback to teachers and as a tool to help districts improve instruction (Kane & Staiger, 2012). We have selected this dataset for secondary data analysis for several reasons. First, MET LDB contains an exceptionally large number of records of practice including data on its instructional quality gleaned from the most well developed instruments to-date. Second, the MET school districts and measures are well linked to the fundamental problems that Mann and Ball argue for studying beginning teaching. In particular, the MET project focused on schools from urban school districts that have high proportions of high poverty and high minority student populations, and their researchers found links between measures of teaching quality and student achievement. Given our research teams' concern for addressing pressing issues with educational equity, the use of MET LDB allows us to consider differences in teaching practice during instruction on mathematical content inside a space where concerns for educational equity are potentially linked to the fundamental problems of beginning mathematics teaching. Therefore, this dataset has promise for secondary analysis of teaching and investigating beginning and experienced mathematics teaching. For this reason, we ask, is there evidence of differences in instructional quality between beginning and experienced teachers during mathematics instruction and what is the nature of those differences? Additionally, the National Science Foundation is encouraging educational research to use existing data like the MET LDB for studying other pressing education issues. We have undertaken such a use. Thus, we also ask, how well has this dataset served our needs?

Methods

Analytic Sample

This research uses the MET LDB collected from approximately 500 volunteer fourth and fifth grade teachers from five urban school districts who video-recorded, on average, two to four self-selected mathematics lessons per year during the 2009-2010 and 2010-2011 school years. Researchers trained on the evaluation instruments scored a sample of these videotaped lessons using six observation protocols. Our study uses the data gleaned from the mathematics lessons using the MQI, CLASS, and FFT instruments. We also draw on scores of teachers' mathematical knowledge for teaching (MKT) as well as information on student performance, lesson context, and teacher and student demographic information.

Items and Scoring

For more information about MQI, CLASS, and FFT items and scoring, please see *User Guide to Measures of Effective Teaching Longitudinal Database* and "Measures of Effective Teaching Longitudinal Database: Documentation" (ICPSR, retrieved on 2016; University of

Michigan, retrieved on 2016).

Defining beginning teaching. We define “beginning teachers” as those in their first through third year of experience in the first year of the MET study. All other teachers are considered “experienced teachers”. Our realized sample includes approximately 60 (or 12%) teachers classified as beginning.

Analytic Models

We undertook two analytical approaches. In our first approach, three-level hierarchical measurement (item response theory) models were used to explore differences between beginning and experienced teaching along different latent dimensions of teaching measured by multiple elements of an observation protocol (See Raudenbush and Bryk, 2002, Chapter 11 for a technical discussion of this model). Our second approach studies variation in individual MQI, CLASS, and FFT elements. Similar to the measurement models, we fit two-level logistic hierarchical models. In both approaches, the number of levels in the models were determined by the sampling structure, assessment of variance, and measures of reliability. We first fit unconditional models and then fit conditional models including a measure of beginning teaching.

Results

Overall, the CLASS and FFT measures of classroom management pick up some differences between beginning and experienced teaching during mathematics instruction, but on most of the dimensions and elements of MQI, CLASS, and FFT, we did not find significant differences in the teaching of beginning and non-beginners.

Measurement models

From the hierarchical measurement analysis, we found some evidence of significant differences between beginning and experienced teachers during mathematics instruction on MQI, CLASS, and FFT favoring experienced teachers, even when controlling for lesson context, teacher and student characteristics, and mathematical knowledge for teaching. Both CLASS and FFT contain latent measures of classrooms – classroom management scale from CLASS and classroom environment scale from FFT. We found patterns of marginal differences between beginning and experienced teachers at the overall domain level for these latent measures of classroom management. For other latent measures, the CLASS emotional and instructional support scales and the FFT instruction scale, no differences were found.

Element-wise Models

Preliminary results from measurement models for CLASS and FFT domains show that there is a moderate to strong correlation between sub-elements. We found significant differences between beginning and experienced teaching on the "content understanding" element from the CLASS instructional support scale and "managing classroom procedures" element from the FFT classroom management scale favoring experienced teachers. Two elements have marginally significant results, namely, instructional learning format from the CLASS classroom management scale and managing student behavior from the FFT classroom environment scale. Results for MQI elements suggest significant differences for only one element, namely "working with students and mathematics," favoring experienced teachers.

Discussion

Our study was not only interested in substantive results but also considered the usefulness of the MET LDB for our study and secondary data analysis generally. In doing so, we sought to learn what features of a data set promote the study of teaching at scale for both primary and secondary analysis. Several lessons emerged as we worked. First, we learned that the MET LDB has many strengths that enabled our study of a slice of teaching. In particular, the MET project collected data on many teachers and students across multiple regions of the United States and sought to create a sample of teachers that was reasonably similar to the teaching population at large. Additionally, the MET LDB includes background information on teachers and students that aid in secondary analysis. Both the sample of teachers and students and their background information facilitate others' use of the data and situate findings that are more widely applicable. Finally, perhaps the primary strength of the this data set is the large number of records of teaching practice, in this case video-taped lessons, that were collected and rigorously scored using six prominent, researched-based observation protocols. In terms of secondary analysis, these scores allow researchers to investigate questions about teaching, as we have done, without using resources to code videos of lessons, which can be both time consuming and expensive.

Several other lessons from our work concern ways that future studies of teaching, aimed to be useful for secondary analysis, might improve upon MET's efforts. One such lesson is the need for diligence in the collection of background information on study participants even when such information is not the focus of the primary study. Our study was limited by missing information on the years of experience measure. Twenty percent of teachers in the sample we drew from the MET LDB were missing years of experience information. We recognize that education research in general tolerates a fairly high percent of missing data. From our work with this data, however, we found this rate of missing data to be troublesome. First, we are perplexed at how data that is collected and used by districts for many purposes including pay rates can be missing at a rate of 20%. Furthermore, in our ponderings of who might be in this group of missing, we cannot help but wonder if this group might include first year teachers, a group of interest to our study, since their data may not be entered. Also, from descriptive statistics comparisons with national descriptive, we are concerned that this group might include a large number of male and minority status teachers...We continue to keep an understanding of why this might be the case in practice. We are further concerned by statistical results from models where this group has been included. Intriguingly, this group with missing years of experience was negative and significantly related to many of our measures compared to experienced teachers. We continue to ponder what this missing-ness means and what to make of our statistical results, but find that having accurate records on this measure, and other background characteristics, is important to our future studies of beginning teaching and to a dataset of teaching intended for secondary analysis. While ensuring high response rates for all background characteristics, particularly those unrelated to the primary study, may seem too costly and time consuming, such efforts will increase the usefulness of the data for secondary analysis.

Our final lesson pertains to sample size of lessons. We mentioned above the advantage of having a large sample of teachers and students, but throughout our use of the MET LDB, we wrestled the low sampling of lessons for each teacher. In this study, we used multi-level models to account for the inter-dependent nature of schooling. That is, we used multilevel models to account for the ways lessons are nested within classes and other ways that scores of teaching might be interdependent. Accounting for the hierarchical and complicated relationships in schools aids in ensuring unbiased results and is becoming increasingly common in education research. Yet, such methods not only require large samples of lessons, teachers, or students, but also multiple lessons for each teacher. With few lessons for each teacher—as few as two, we found that our models had low reliability and thus, had limited ability to distinguish the true affect of teaching from noise. For this reason, sampling design for studies of teaching should be thoughtful about the optimum number of lessons needed for each teacher or class.

Overall, we still see much potential in using the MET LDB for secondary analysis of instruction, yet as our field moves forward, we can learn from the challenges of using this dataset. We also recognize that the work of designing, collecting, and scoring records of teaching practice is no small endeavor and limits of funding and time often produce tradeoffs and compromises. Yet, we think funding primary studies with an eye toward secondary analysis, as NSF is promoting, in the long run will enable more researchers to investigate questions important to improving instruction in the U.S.

In light of the advantages and limitations of using MET LDB for our secondary study of beginning mathematics teaching, we consider what we learned about beginning mathematics teaching. Across the FFT results, we found meaningful differences between beginning and experienced teachers on classroom environment but no differences on instruction. The classroom environment scale is said to measure "if an environment is conducive to learning" while the instruction scale assesses "students' actual engagement with the content" (ICPSR, retrieved on 2016). These differences seem to suggest that beginning teachers are less able to provide a conducive learning environment while they are as skilled as more experienced teachers in providing a space for students' actual engagement with the content. While we found no differences between beginning and experienced teachers on the measures of instruction, we note that all teachers, beginning and experienced, struggle to perform well on the measures of instruction. Our finding may reflect that rigorous instruction is challenging for teachers across years of experience and not the skill of beginners.

While CLASS and FFT measure similar dimensions of teaching (e.g.. behavior management, emotional support, instruction), we did not find the same patterns with the CLASS scales and items. The primary differences between beginners and non-beginners fell along one of the instructional measures, content understanding. This aspect of CLASS measures the quality of "interactions among the teacher and students that lead to an integrated understanding of facts, skills, concepts, and principles" (ICPSR, retrieved on 2016). Additionally, our analysis of MQI also found differences in the instructional techniques between beginners and more experienced teachers, as beginners performed less well on the "working with students and mathematics" measure.

Altogether our findings suggests that beginning teaching may differ from more experienced teaching on multiple aspects of teaching including behavior and classroom management and some instructional techniques. The many dimensions upon which beginning and experienced teachers did not differ, however, may mean that in many ways beginning and experienced teaching is similar. Yet, the challenges to our analysis discussed above may have limited our ability to detect true difference in teaching between these groups of teachers. Certainly, further research is needed to confirm our findings and explore beginning teaching.

References

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational evaluation and policy analysis, 25(2),* 119-142.

Dunkin, M. J., & Biddle, B. J. (1974). The study of teaching. Holt, Rinehart & Winston.

Farmer, S. O., Mann, L., Hickman, L. A., & Snider, R. B. (2015). Coordination in beginning mathematics teaching. In S.M. Wilson (Chair), Investigating beginning mathematics teaching. Symposium conducted at the annual meeting of the National Council of Teachers of Mathematics, Boston, MA.

Gordon, R. J., Kane, T. J., & Staiger, D. (2006). Identifying effective teachers using performance on the job. Washington, DC: Brookings Institution.

Harrison, D. M., Gadd, K. J., Johnson, L. J., & Mann, L. (2015). Challenges to studying beginning mathematics teaching at scale. In S.M. Wilson (Chair), Investigating beginning mathematics teaching. Symposium conducted at the annual meeting of the National Council of Teachers of Mathematics, Boston, MA.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. Second handbook of research on mathematics teaching and learning, 1, 371-404.

ICPSR. (Retrieved on March 13, 2016). Measures of Effective Teaching Longitudinal Database: Documentation. Retrieved from http://www.icpsr.umich.edu/icpsrweb/METLDB/holdings/documentation.

Ingersoll, R., & Merrill, L. (2010). Who's teaching our children. Educational Leadership.

University of Michigan, Ann Arbor, Inter-University Consortium for Political and Social Research. (2014, December). User Guide to Measures of Effective Teaching Longitudinal Database (MET LDB). Retrieved July 22, 2015, from http://www.icpsr.umich.edu/icpsrweb/METLDB/studies/34771.

Kane, Thomas J., Staiger, Douglas O. Jan 2012. Gathering Feedback for Teaching: Combining High-quality Observations with Student Surveys and Achievement Gains. MET Project Research Paper. Bill & Melinda Gates Foundation

Lampert, M. (2001). Teaching problems and the problems of teaching. New Haven, CT: Yale University Press.

Lotie, D. C. (2002). Schoolteacher: A sociological study. (2nd Edition). The University of Chicago Press, Chicago.

Mann, L. & Ball, D. L. (2015). Why Study Beginning Teaching? In S.M. Wilson (Chair), Investigating beginning mathematics teaching. Symposium conducted at the annual meeting of the National Council of Teachers of Mathematics, Boston, MA.

Monk, D. H., & King, J. A. (1994). Multilevel teacher resource effects in pupil performance in secondary mathematics and science. The case of teacher subject matter preparation. Choices and consequences: Contemporary policy issues in education, 29-58.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. (Vol. 1). Sage.

Richardson, V. (Ed.). (2001). Handbook of research on teaching. American Educational Research Association.

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.

Vanneman, A., Hamilton, L., Baldwin Anderson, J., and Rahman, T. (2009). Achievement Gaps: How Black and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress, (NCES 2009-455). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.