

6-2017


Relationships Between Mathematics Teacher Preparation and Graduates' Analyses of Classroom Teaching

James Hiebert
University of Delaware

Emily Miller
West Chester University of Pennsylvania, emiller@wcupa.edu

Dawn Berk
University of Delaware

Follow this and additional works at: https://digitalcommons.wcupa.edu/math_facpub

 Part of the [Science and Mathematics Education Commons](#), and the [Teacher Education and Professional Development Commons](#)

Recommended Citation

Hiebert, J., Miller, E., & Berk, D. (2017). Relationships Between Mathematics Teacher Preparation and Graduates' Analyses of Classroom Teaching. *Elementary School Journal*, 117(4), 687-707. <http://dx.doi.org/10.1086/691685>

This Article is brought to you for free and open access by the College of the Sciences & Mathematics at Digital Commons @ West Chester University. It has been accepted for inclusion in Mathematics by an authorized administrator of Digital Commons @ West Chester University. For more information, please contact wccressler@wcupa.edu.

RELATIONSHIPS BETWEEN MATHEMATICS TEACHER PREPARATION AND GRADUATES' ANALYSES OF CLASSROOM TEACHING

ABSTRACT

The purpose of this longitudinal study was to investigate the relationships between mathematics teacher preparation and graduates' analyses of classroom teaching. Fifty-three graduates from an elementary teacher preparation program completed 4 video-based, analysis-of-teaching tasks in the semester before graduation and then in each of the 3 summers following graduation. Participants performed significantly better on the 3 tasks focused on mathematics topics studied in the program than on the task focused on a mathematics topic not studied in the program. After checking several alternative hypotheses, we conclude that a likely explanation for the performance differences is the mathematical knowledge for teaching that participants developed as freshmen in the preparation program.

James Hiebert
UNIVERSITY OF
DELAWARE

Emily Miller
WEST CHESTER
UNIVERSITY

Dawn Berk
UNIVERSITY OF
DELAWARE

THE major question facing teacher educators today is whether and how their teacher preparation programs matter. The United States has little information about how its programs make a difference (Borko & Putnam, 1996; Cochran-Smith & Zeichner, 2005; Feuer, Floden, Chudowsky, & Ahn, 2013; Levine, 2006; National Research Council, 2010; Pajares, 1992; Zeichner & Tabachnick, 1981). Vague descriptions of what actually occurs in preparation programs, together with few systematic efforts to follow graduates into the field, have left teacher educators guessing about the ingredients to include in effective programs.

THE ELEMENTARY SCHOOL JOURNAL
Volume 117, Number 4. Published online May 3, 2017
© 2017 by The University of Chicago. All rights reserved. 0013-5984/2017/11704-0007\$10.00

This study followed graduates of an elementary teacher preparation program 3 years beyond graduation and tracked their performance on an analysis-of-teaching task, a version of which has been shown to correlate with mathematics teaching quality and student learning (Kersting, Givvin, Thompson, Santagata, & Stigler, 2012). The question is whether the mathematics studied during the preparation program relates to graduates' performance on this task up to 3 years after graduation. In what ways might teacher preparation have made a difference?

Relationships between Teacher Preparation and Quality of Teaching

Despite years of research, there is little evidence documenting the effects of teacher preparation on the quality of teaching (Arbaugh, Ball, Grossman, Heller, & Monk, 2015; Cochran-Smith et al., 2015; Cochran-Smith & Zeichner, 2005). Although researchers have found some programs more effective than others (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Gansle, Noell, & Burns, 2012; Lincove, Osborne, Mills, & Bellows, 2015), it has been difficult to explain what features of teacher preparation make a difference (Cochran-Smith et al., 2015; Diez, 2010). In general, clinical experiences during teacher preparation seem to have more positive effects than content preparation on graduates' teaching quality (Arbaugh et al., 2015; Cochran-Smith et al., 2015; Henry et al., 2013).

The effects of content knowledge acquired during teacher preparation have been mixed and relatively weak. Monk (1994), for example, found the number of courses taken by preservice mathematics and science teachers predicted their students' learning, but the relationship was nonlinear. In some cases, beyond a threshold level, more mathematics courses had diminishing returns. In their extensive review, Floden and Meniketti (2005) concluded that, overall, only small correlations have been found between subject matter coursework taken during a preparation program and graduates' performance in the classroom.

Our goal in this study was to explore relationships between the mathematics preservice teachers (PSTs) studied during their preparation program and their performance when analyzing videos of classroom teaching of the same mathematics. Given the challenge of finding relationships between teacher preparation and teaching quality (Diez, 2010), we reasoned that using specific measures of teaching-like skills, aligned with the specific kinds of knowledge that PSTs had opportunities to develop during preparation, would increase our chances of documenting significant relationships (Lincove et al., 2015).

Integrating Three Lines of Research

Performance on mathematics analysis-of-teaching tasks brings together three lines of research on competencies that relate to the quality of mathematics teaching.

Mathematical Knowledge for Teaching

One line of work that provides a foundation for this study centers around the construct of mathematical knowledge for teaching (MKT; Ball, Thames, & Phelps,

2008), an elaboration and specification of Shulman's (1986) pedagogical content knowledge for school mathematics. MKT is related to, but different from, conventional knowledge of mathematics. By definition, it is the kind of knowledge that teachers use when they teach mathematics. The elements of MKT were derived from analyzing the work of teaching mathematics and include the knowledge needed (1) to select and use appropriate representations or models of mathematical ideas and procedures, (2) to interpret students' responses and evaluate the appropriateness of their solution strategies, and (3) to connect mathematical ideas being studied with those students already know (Ball et al., 2008; Hill et al., 2008).

The usefulness of MKT for studying teacher preparation comes, in part, from the fact that it is precise enough to capture the knowledge teachers are likely to use when teaching, but broad enough to apply across a range of mathematics topics. Empirically, the relation between MKT and mathematics teaching and student learning has been mixed. Those who initially defined MKT have reported significant relationships with quality of mathematics teaching and, in turn, with students' learning (Ball & Bass, 2000; Ball et al., 2008; Hill et al., 2008; Hill, Rowan, & Ball, 2005; Hill, Sleep, Lewis, & Ball, 2007). However, others have found fewer significant relationships with both the quality of teaching and students' learning (Kersting, Givvin, Sotelo, & Stigler, 2010; Kersting et al., 2012; Ottmar, Rimm-Kaufman, Larsen, & Berry, 2015). Although the empirical relation among MKT, teaching, and student learning still must be clarified, there are strong theoretical arguments that link MKT with the knowledge mathematics teachers need.

Teacher Noticing

MKT is an important component of teachers' competence, but it is far from sufficient. Researchers argue that another important competency is being able to examine teaching in particular ways. This leads to a second line of work relevant to our study—teacher noticing. The rationale underlying this work is that teachers must *see* teaching differently before they can change the way they teach (Sherin, Jacobs, & Philipp, 2011; van Es & Sherin, 2002). This follows from the claim that what teachers notice is associated with what they do (Erickson, 2011; Jacobs, Lamb, Philipp, & Schappelle, 2011; Schoenfeld, 2011). For example, if teachers viewing a lesson attend to student behavior, classroom management, or the way students are grouped for instruction, then they are likely to focus on these same features when teaching their own students. Alternatively, if they notice the mathematics being discussed and the ways students think about it, they are likely to focus on these features in their own teaching. It is reasonable to assume that teachers with more deeply developed MKT would be more inclined, and more able, to attend to the mathematics in classroom interactions.

The work on teacher noticing reveals a developmental sequence in novice teachers (Santagata & Angelici, 2010; Santagata, Zannoni, & Stigler, 2007; Sherin & van Es, 2005; Star & Strickland, 2008; Stockero, 2008; van Es & Sherin, 2008). When asked to describe a classroom episode, they initially provide descriptive or evaluative responses, simply describing what they see and often focusing on surface features. Influenced by guided experiences, they move to more interpretative and analytic responses that lead to cause-effect kinds of reasoning. More advanced novices begin

to propose hypotheses for how teachers' actions lead to particular student responses, a form of reasoning associated with improvement of teaching that supports richer student learning (Gallimore, Ermeling, Saunders, & Goldenberg, 2009).

Analysis of Teaching

The third line of research relevant to this study is the development of an assessment that measures teachers' tendency to notice particular aspects of mathematics teaching. This assessment has been shown to predict the quality of teaching and, in turn, the nature of student learning (Kersting, 2008; Kersting et al., 2010, 2012). The assessment consists of watching a video clip of actual classroom instruction and describing what happens. Kersting and colleagues scored teachers' responses in terms of what they attended to: mathematics content, students' thinking, alternative methods the teacher could have used, and overall quality of the classroom interaction. The third category, proposing alternative methods, was found to be most highly correlated with the quality of the participants' teaching and their students' learning. Proposing alternative methods of interacting with students about mathematics would seem to require both deep MKT of the topic and advanced noticing skills.

Our review of the literature indicates that the analysis-of-teaching video task, developed and tested by Kersting and colleagues (2010, 2012), has strong theoretical and empirical links to higher quality mathematics teaching and increased student learning. The theoretical arguments supporting the importance of what teachers notice, and the empirical data linking the nature of teachers' responses to the video tasks to both the quality of their teaching and their students' learning, suggest that analyzing videos of teaching assesses important teaching-related skills. The video task also has a practical advantage—it is administrable online and thus enables the assessment of many more teachers, across multiple years, than would be possible with on-site observations.

Setting for the Study

In this article, we report performance data on an analysis-of-teaching video task as part of a longitudinal study of the effects of mathematics teacher preparation. The teacher preparation of interest is the mathematics portion of an elementary teacher preparation program housed in the School of Education at the University of Delaware. This is a 4-year certification program in which all students earn K–6 certification, and many choose a second certification in a middle school subject area (for grades 6–8). The program graduates about 130 students per year. All students in the program complete three mathematics content courses and one elementary mathematics methods course. Multiple sections of these courses are offered each semester; all sections are taught by mathematics education faculty and doctoral students.

The three mathematics content courses develop key mathematical ideas from the K–6 curriculum. Much of the material is encountered through activities that simulate authentic teaching tasks to develop PSTs' MKT. These tasks include an-

alyzing student work, critiquing students' explanations and creating explanations students can understand (by connecting new ideas with those students already know), creating visual representations for key mathematical concepts and procedures, and modeling solution strategies students might use. There is a clear parallel between the nature of these instructional tasks and the knowledge components of MKT specified earlier. In addition, the courses emphasize two pedagogical moves shown to support students' conceptual understanding: making the key mathematical ideas explicit during instruction and providing students with opportunities to productively struggle with these ideas (Hiebert & Grouws, 2007).

This study focused on the MKT developed during the first two of three content courses. Most PSTs take these courses as freshmen. The first course covers whole number and decimal numeration systems. The course includes a study of ancient numeration systems, place value concepts, counting and measuring quantities represented with whole numbers and decimals, and the arithmetic operations on whole numbers and decimals. Arithmetic operations are developed through writing story problems, concretely modeling decimal number problems to create meaning for the operations, analyzing children's solution strategies, and examining the conceptual basis for standard and nonstandard algorithms for operating on whole numbers and decimals.

The second content course focuses on fractions and proportional reasoning. This course involves a detailed study of different meanings for and representations of fractions, the arithmetic operations on fractions, and different representations and methods for solving proportional reasoning problems. As in the first course, arithmetic operations on fractions are developed through writing story problems, concretely and visually modeling both fractions and operations on fractions (e.g., modeling subtraction of fractions using fraction strips), analyzing children's solutions to fraction and proportional reasoning problems, and unpacking the conceptual basis for standard and nonstandard algorithms for operating on fractions.

The third content course and the elementary mathematics methods course are not implicated directly in the tasks used in this study. The third content course focuses mostly on geometry, with a few lessons devoted to measurement and algebra. The methods course focuses on pedagogical issues and includes an intensive field experience. The mathematics topics of interest in this study were not treated extensively in the methods course. Most PSTs take the third content course when they are sophomores and the methods course when they are juniors or seniors. In none of the four courses did PSTs complete tasks exactly like those administered in this study.

A unique and important design feature of this program is that all PSTs receive nearly the same instruction regardless of which course section they take. Over the past 15 years, our mathematics education group has created detailed lesson plans for each session of each content course (Berk & Hiebert, 2009; Hiebert & Morris, 2009). Instructors of each course meet weekly to discuss the lesson plans and their enactment. As a result, all instructors who teach these courses provide PSTs with the same instructional activities. In addition, all instructors use common homework assignments and common exams each semester. Although we cannot verify that the instructional activities were implemented in exactly the same way by all instructors, we can claim that students receive very similar learning opportunities.

One way to test the claim of similar learning opportunities across sections of the same course is to compare final exam scores across sections. All PSTs taking the same course complete a common final exam. A one-way between-subjects ANOVA revealed no statistically significant differences between section means for the final exam in the first content course (Fall 2007), $F(5, 152) = 1.20, p = .31$. In the second content course (Spring 2008), the difference between section means approached significance, $F(5, 125) = 2.34, p = .05$. However, post hoc comparisons using the Tukey HSD test showed no significant differences between any pair of sections (all $p > .05$). These results lend support to the claim that PSTs in different sections experienced similar learning opportunities.

Method

Sample

All students graduating from the program in 2011 were invited to participate in the study. Of the 131 graduates, 97 initially volunteered. Over the course of the next 3 years, attrition resulted in a final sample of 53 graduates who completed the tasks each year. These 53 graduates compose the sample for this study.

The 53 participants achieved a combined grade-point average of 2.98 (4.00 maximum) in the two mathematics content courses most relevant for this study, compared with a combined grade-point average of 2.70 for the 78 nonparticipants. This difference is significant ($U = 1,651, p = .045$). Consequently, the participants can be considered slightly more mathematically prepared than their nonparticipating peers.

All participants were paid a stipend for completing a larger collection of tasks, of which the analysis-of-teaching tasks were a subset. Of the 53 participants, the following numbers were regular classroom teachers in each of their first 3 years after graduation: first year, 38; second year, 45; third year, 45. The remaining participants were attending graduate school, employed in nonteaching positions, or not currently employed.

Research Design

One dilemma for researchers investigating the effects of teacher preparation on graduates' competencies is whether to include a control group. Despite the advantages of doing so, comparing the performance of graduates of one program with graduates of another program does little to isolate the effects of the program of interest because the differences among programs in goals and experiences are likely to be numerous and complicated. This makes it nearly impossible to create a reasonable comparative measure of effects and to explain the source of the differences that might be found. Consequently, such comparisons would provide little insight into whether and how teacher preparation matters for acquiring teaching-related skills and practices.

Because a control group would not help us interpret the connections between what PSTs studied in the program and their use of this knowledge after they graduated, we designed this study so each graduate would serve as her or his own con-

trol. Graduates analyzed video clips focused on four different mathematics topics. Three target topics—multiplying two-digit whole numbers, subtracting fractions, and dividing fractions—were topics that received considerable attention in the first two mathematics content courses. The control topic—finding the mean for a small set of whole numbers—is a topic not covered at any point in the program. All four topics receive attention in current U.S. elementary school curricula, and each has an associated standard algorithm whose meaning derives from several underlying concepts. We hypothesized that relationships between the teacher preparation program and graduates' analyses of teaching should emerge as higher quality analyses of the video clips on the target topics than on the control topic.

Tasks. A brief video clip from a classroom lesson on each of the four mathematics topics was selected (see Kersting, 2008). The clips ranged from 4 to 6 minutes in duration. Each clip showed a teacher interacting with students around a mathematics problem involving the topic of interest. Each video clip exhibited some deficiencies in the interaction that could limit students' opportunities to understand the mathematics. In other words, the clips were intentionally chosen so participants could legitimately critique the conceptual opportunities provided by the instructional interaction and propose an alternative instructional strategy. The task for each topic asked participants to respond to the following prompt: "View the clip and discuss how the teacher and the student(s) interact around the mathematical content." This is the same prompt used by Kersting and colleagues (2010). No length requirements for the responses were suggested or required.

The video clip for multiplying two-digit whole numbers showed a third-grade teacher lead students step by step through the standard algorithm for solving 52×36 . Using a recitation form, the teacher asked students short-answer and fill-in-the-blank questions as she moved through the procedure. One student asked the teacher why she put a zero in the one's place before beginning to multiply by the three 10s in 36. The teacher said the zero was a placeholder and would prevent students from putting another number in that place. Throughout the interaction, the teacher complimented the students and exuded enthusiasm, which seemed to keep them engaged.

The video clip for subtracting fractions began with a fifth-grade teacher asking students to put out 12 blocks and then take away 3. The teacher then said, "From what remains, subtract $\frac{1}{3}$." Students worked for a short time and then began offering answers. Some students said the answer was 6, some said $\frac{6}{12}$, and some said $\frac{5}{12}$. The teacher tried to clarify the problem by asking how many blocks would represent $\frac{1}{3}$. She then restated the problem as "We have $\frac{9}{12}$, now take away $\frac{1}{3}$ of what's left." The confusion about what counted as the referent for $\frac{1}{3}$ was never resolved.

The video clip for dividing fractions showed a sixth-grade teacher soliciting student solutions to the problem $\frac{1}{2}$ divided by $\frac{2}{3}$. A student said the answer was $1\frac{1}{6}$. The teacher asked him to come to the board and "show how $1\frac{1}{6}$ of your $\frac{2}{3}$ go inside $\frac{1}{2}$." The student began to present his solution using the circular fraction pieces on the board, but the teacher quickly interrupted him. The teacher said they were trying to find how many $\frac{2}{3}$ s go into $\frac{1}{2}$. Another student suggested the answer was $\frac{5}{6}$, to which the teacher responded "Close, but look at it here." The teacher then modeled the problem at the board using the fraction pieces, and

asked, “If $\frac{2}{3}$ is my whole, then how much of this whole fits inside this $\frac{1}{2}$ piece?” A third student responded “ $\frac{3}{4}$ ” and the teacher confirmed this as the correct answer. The teacher then showed students how $\frac{3}{4}$ of the $\frac{2}{3}$ piece fit into the $\frac{1}{2}$ piece. In the clip, the teacher interacted with students in a friendly, engaging way but did the bulk of the mathematical work for the students.

In the video clip for finding the mean, a fourth-grade teacher asked students to find the number of pets each of seven families could have if the mean number of pets was 4, but no family had exactly four pets. The teacher asked students to work in groups and circulated around the classroom, giving hints when needed. The teacher helped a pair of students see they could begin by counting seven 4s to find the total number of pets. Later, the teacher asked the same students how many pets there would be altogether and then asked how this number could be distributed among the seven families. Finally, the teacher suggested giving each family four pets and then redistributing them. As with the dividing fractions video clip, the teacher interacted in a pleasant, engaging way with the students but did most of the mathematical work for them.

Participants first completed the four analysis-of-teaching tasks in the semester before graduation (Spring 2011) and then during the summers following each of the next 3 years (Summer 2012, 2013, and 2014). The analysis-of-teaching tasks were posted online, accessible via passcode. Participants were directed to click on a link to view the video clip, and then they typed their response to the prompt into the online system.

The tasks were presented in the following order: multiplying two-digit whole numbers, subtracting fractions, finding the mean, and dividing fractions. However, because no order was prescribed and because the tasks were presented online, participants could complete these tasks in any order they chose, and they could complete the tasks across multiple sittings (e.g., they could log in to the system, complete one or two tasks, log out, and then return at a later time to complete the remaining tasks). These features of the system allowed us to confirm that many participants completed the tasks in orders different than the order presented. For example, some participants completed the subtraction of fractions task first, whereas others completed this task last. We were not able, however, to determine the exact order in which every participant completed the tasks. So, although we cannot rule out an order effect, this seems an unlikely explanation for our findings.

To track changes over time, participants analyzed the same video clips each year. To minimize the potential bias of having participants submit identical responses each year, we designed the online system so participants were unable to access the videos during the year or read their responses from previous years. No feedback was provided to the participants on their responses. We also verified that participants’ responses were not identical from year to year.

Coding. Responses were coded using a rubric modified from that used by Kersting and colleagues (2010, 2012). We drew primarily from three sources to develop our rubric: (1) the developmental progression from describing the interaction to analyzing and critiquing the interaction identified in the “teacher noticing” literature; (2) Kersting and colleagues’ (2010) finding that proposing alternative instructional strategies was the dimension most highly associated with quality teaching; and (3) Hiebert and Grouws’s (2007) claim that students’ conceptual understanding is

supported by two pedagogical moves—allowing students to productively struggle with the mathematics and making the key mathematical ideas in the lesson explicit for students.

Integrating these factors, we created two scales: one for noticing the mathematics in the video episode and one for critiquing the pedagogy and the mathematical interactions in the video episode. For each video clip response, a mathematics score and a critique score were assigned, each ranging from 0 to 4. The mathematics score was created by adding two subscores, each ranging from 0 to 2. One subscore assessed the degree to which the participant described the mathematics in the video; the other subscore assessed the degree to which the participant critiqued the mathematics and offered suggestions for adjusting the mathematics to facilitate better conceptual understanding of the topic. The critique score was also created by adding two subscores, each ranging from 0 to 2. One subscore was the subscore used for mathematics critique (just described); the other subscore assessed the degree to which the participant critiqued the pedagogy and offered suggestions aligned with the two pedagogical moves identified earlier for supporting students' conceptual understanding. Table 1 shows the definitions for each subscore in the rubric as well as the key mathematical idea for each topic. Examples of responses, along with detailed explanations of scoring responses, are available from the authors.

Note that the mathematics critique subscore is part of both the mathematics score and the critique score. This reflects the fact that these constructs are not independent. At the most advanced level, noticing the mathematics involves cause-effect reasoning that suggests changes, where appropriate, to the mathematics discussed in the classroom to improve the conceptual learning opportunities for students. Critiquing the video episode includes critiquing the mathematics as well, in

Table 1. Rubric for Coding Responses to Analysis-of-Teaching Tasks

Code	Score		
	0	1	2
Mathematics, descriptive	Identified the mathematics with, at most, one word or label	Described the mathematics with at least a phrase but did not identify the key mathematical idea in the episode	Described the key mathematical idea in the episode ^a
Mathematics, critique	Offered no suggestions for clarifying or interacting about the mathematics	Suggested something could have been done differently (e.g., "used a different problem") but did not address the key idea	Suggested the key idea should have been addressed differently and offered suggestions for how to do this
Pedagogy, critique	Offered no critique of the pedagogy and made no suggestions for how the pedagogy could have been improved ^b	Suggested the pedagogy could have been improved but did not reference productive struggle or making mathematics ideas explicit	Suggested improving the pedagogy by providing specific opportunities for students to productively struggle or by making key ideas explicit

^a The key mathematical ideas for each topic follow: *multiply whole numbers*—zero is used to "hold a place" because the 6×5 in the problem is 6×50 , so the result is 300, not 30; *subtracting fractions*—identify the materials that count as one (or count as one of the key unit fractions in the problem); *dividing fractions*—two thirds fits into one half less than one time, so two thirds must be partitioned to make part of it fit; *finding the mean*—the total for a set of numbers is the number of items times the average size of each item *even though* each item has more or less than the average.

^b All video clips were instructionally deficient in ways that invited suggesting alternative pedagogical moves.

just the same way. Consequently, we created mathematics and critique scores that both included the mathematics critique subscore.

Responses were coded by the first author. Interrater reliability was assessed by asking a second coder (a faculty member in mathematics education) to code responses from the first year of data collection from 10% (6) of the 53 participants chosen randomly. Reliability was computed for each code by comparing the number of agreements with the total number of decisions (6 participants \times 4 topics = 24 decisions per code). Rate of agreement was 96.0% for pedagogical critique, 87.5% for mathematics descriptive, and 96.0% for mathematics critique.

Because the participants are graduates of a teacher preparation program in which we teach, it is important to understand our relationship to the participants and how this might have affected the results. The first and second authors taught none of the participants during their program; the third author might have taught some of the participants in the first and/or second content course. The first author did not know the identity of the participants while coding the data because the participants' responses were blinded by project staff and assigned code numbers. Communication with participants about completing this task was handled by the third author.

Although we included these safeguards against knowingly biasing the results, we were aware of which topics were studied in the program and which were not. This knowledge could have biased our coding. However, one additional consideration is relevant. Because we established interrater reliability, the most likely source of potential bias would be built into the coding rubric. We could have unconsciously created coding rubrics that favored the target topics over the control topic. However, we actively tried to create parallel rubrics, and we have made the rubrics available for inspection (see Table 1).

Analyses. To compare the effect of topic, as well as time, two-way repeated-measures ANOVAs were conducted for the total mathematics scores (0–4) and the total critique scores (0–4). Each analysis included two within-subjects factors, each with four levels: topic (multiplication, subtraction, division, and mean) and time (in years since graduation, ranging from 0 to 3). Post hoc tests for significant interactions were conducted by calculating simple main effects using a Bonferroni correction for multiple comparisons. To maintain conventions for critical p values, p values resulting from the post hoc tests were multiplied by the number of comparisons being conducted (six in all cases) and then compared with conventional critical values to determine significance. Mauchly's test of sphericity was used to test the equality of variance assumption of the two-way repeated-measures ANOVA. A significant test could indicate an inflated F statistic and an increased Type I error rate; in these cases, the Greenhouse-Geisser correction was applied to attenuate the F statistic.

Results

Mathematics Scores

Means and standard deviations of participants' total mathematics scores for the four mathematics topics in Years 0 to 3 are shown in Table 2. Mauchly's test of

Table 2. Total Mathematics Scores (Maximum of 4 Points)

Topic	Years since Graduation, <i>M</i> (<i>SD</i>)			
	0	1	2	3
Multiplication	.81 (.52)	.98 (.93)	1.43 (1.01)	1.47 (1.46)
Subtraction	.93 (1.09)	1.43 (1.19)	2.02 (1.34)	1.45 (1.19)
Division	.66 (.76)	1.21 (.99)	1.40 (.93)	1.21 (.95)
Mean	.55 (.54)	.94 (.69)	1.25 (.94)	.89 (.87)

sphericity indicated the assumption of sphericity was not violated for the main effect for time but was violated for the main effect for topic and for the interaction term ($p = .013$, $p < .001$, respectively). Therefore, the F statistics for these two effects were corrected using the Greenhouse-Geisser correction. There was a significant main effect for time, $F(3, 156) = 22.809$, $p < .001$, as well as a significant main effect for topic, $F(2.524, 131.233) = 12.807$, $p < .001$. There was also a significant interaction between time and topic, $F(5.904, 307.023) = 2.190$, $p = .045$, indicating that the effect of topic depends on time and the effect of time depends on topic.

For the mathematics scores, a test of the simple main effect for topic revealed significant differences between target topics and the control topic in each of the 4 years of the study. In each case, when significant differences were found between a target topic and the control topic, they favored the target topic. At graduation, the difference in performance was significant between multiplication and the mean ($p = .012$). At 1 and 2 years after graduation, performance on subtraction was significantly better than performance on the mean ($p = .024$ and $p < .001$, respectively). At 3 years after graduation, performance on all three target topics was significantly better than performance on the control topic (multiplication, $p = .012$; subtraction, $p = .008$; division, $p = .032$).

A test of the simple main effect for time for the mathematics scores indicated that the scores on all three target topics showed significant improvement when comparing participants' initial scores at graduation with their scores 3 years after graduation (multiplication, $p = .007$; subtraction, $p = .030$; division, $p = .001$). However, participants' scores on the mean 3 years after graduation were not significantly higher than their initial scores. Although the participants' multiplication scores did not decrease during any year of the study, their subtraction, division, and mean scores appeared to decrease in the last year. This decrease was statistically significant for subtraction and the mean ($p = .005$ and $p = .039$, respectively), but not for division.

A graph of the changes in average mathematics scores across years is presented in Figure 1. The graph shows that participants' performance on the mathematical analysis for the four topics began at similar places (Year 0, at graduation). Over time, performance on the three target topics moved to significantly higher levels, whereas performance on the control topic did not.

Critique Scores

Means and standard deviations of participants' total critique scores for the four mathematics topics in Years 0–3 are shown in Table 3. Mauchly's test of sphericity

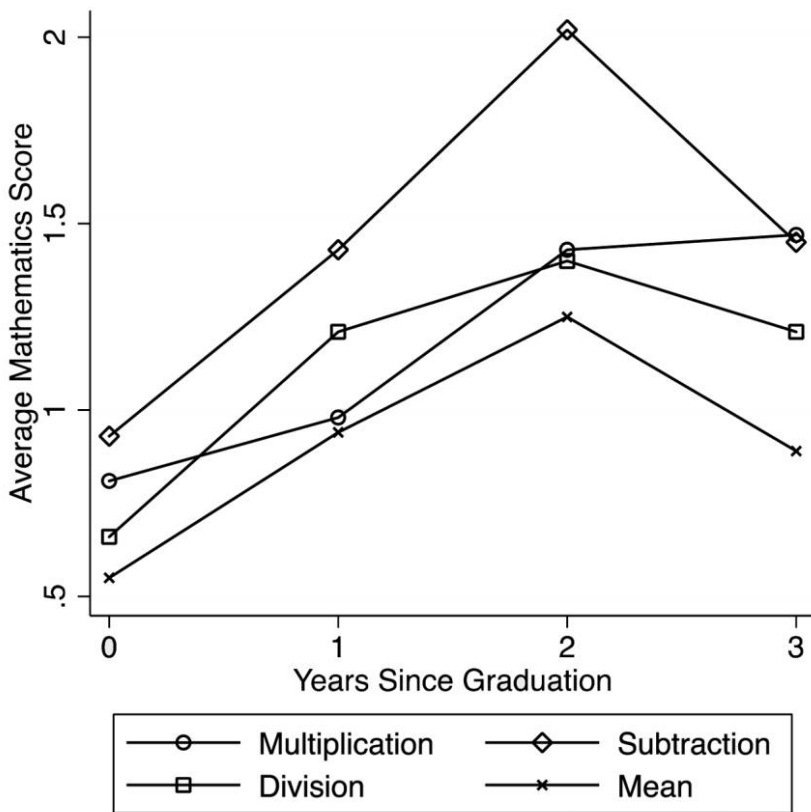


Figure 1. Changes in average mathematics scores for the four topics over time.

indicated that the assumption of sphericity was not violated for either of the main effects but was violated for the interaction term ($p < .001$). Therefore, the F statistic for the interaction term was corrected using the Greenhouse-Geisser correction. There was a significant main effect for time, $F(3, 156) = 11.487, p < .001$, as well as a significant main effect for topic, $F(3, 156) = 16.791, p < .001$. There was also a significant interaction between time and topic, $F(5.602, 291.329) = 4.997, p < .001$, indicating that the effect of topic depends on time and the effect of time depends on topic.

For the critique scores, a test of the simple main effect for topic revealed significant differences between target topics and the control topic at 1, 2, and 3 years after

Table 3. Total Critique Scores (Maximum of 4 Points)

Topic	Years since Graduation, M (SD)			
	0	1	2	3
Multiplication	.23 (.54)	.42 (.84)	.79 (.89)	.87 (1.29)
Subtraction	.55 (.85)	.72 (.79)	1.06 (.95)	.43 (.69)
Division	.34 (.55)	.51 (.78)	.66 (.62)	.19 (.56)
Mean	.23 (.51)	.17 (.38)	.47 (.70)	.09 (.41)

graduation. In each case where significant differences were detected between a target topic and the control topic, they favored the target topic. One year after graduation, the difference in performance was significant between subtraction and the mean ($p < .001$) and between division and the mean ($p = .001$). Two years after graduation, the difference in performance was significant between subtraction and the mean ($p = .001$). Three years after graduation, performance on multiplication and subtraction was significantly better than performance on the mean ($p < .001$ and $p = .008$, respectively).

A test of the simple main effect for time for the critique scores indicated that the participants' scores for multiplication showed significant improvement when comparing their initial scores at graduation to their scores 3 years after graduation ($p = .006$); however, participants' scores on subtraction and division and the mean at 3 years after graduation were not significantly higher than their initial scores. Although participants' multiplication critique scores did not decrease during any year of the study, their subtraction, division, and mean critique scores exhibited a significant decrease in the last year of the study ($p < .001$, $p = .001$, and $p = .001$, respectively).

A graph of the changes in average critique scores across years is presented in Figure 2. The graph shows a more rapid increase in performance on some of the target topics through Year 2, but then a significant decline in performance in Year 3.

One hypothesis for the decline in some scores from Year 2 to Year 3 is fatigue. Participants had been completing these tasks for 4 consecutive years. Perhaps they became tired of completing these tasks and, knowing this was the final year of the project, might have been less motivated to invest as much effort in composing their responses. We tested this hypothesis by comparing the number of words in the participants' responses each year, assuming that fewer words might indicate less effort. Comparing Year 3 and Year 2 responses, we found significant differences only for dividing fractions, and this difference was relatively weak ($p = .021$). Together with the fact that the decline was significant only for some scores, this finding suggests the decline in scores is not likely attributable to lower effort.

A second hypothesis for the decline in scores from Year 2 to Year 3 is that the knowledge graduates acquired during the preparation program deteriorated or was no longer accessed and applied on this task. We cannot rule out this possibility, but it seems a bit odd that the scores on these tasks increased for the first years and then, in some cases, declined only in the final year of the study. We revisit this hypothesis in the Discussion section.

Confirming Differences between Performance on the Target Topics and the Control Topic

The most consistent finding to this point is that, on average, graduates performed better on the video tasks for the target topics (multiplying two-digit whole numbers, subtracting fractions, and dividing fractions) than for the control topic (finding the mean). The pairwise comparisons were not always significant, but, when they were, they always favored the target topic. To confirm that graduates' underlying competence with multiplying two-digit whole numbers, subtracting fractions, and dividing fractions best explains their better performance on the analysis-

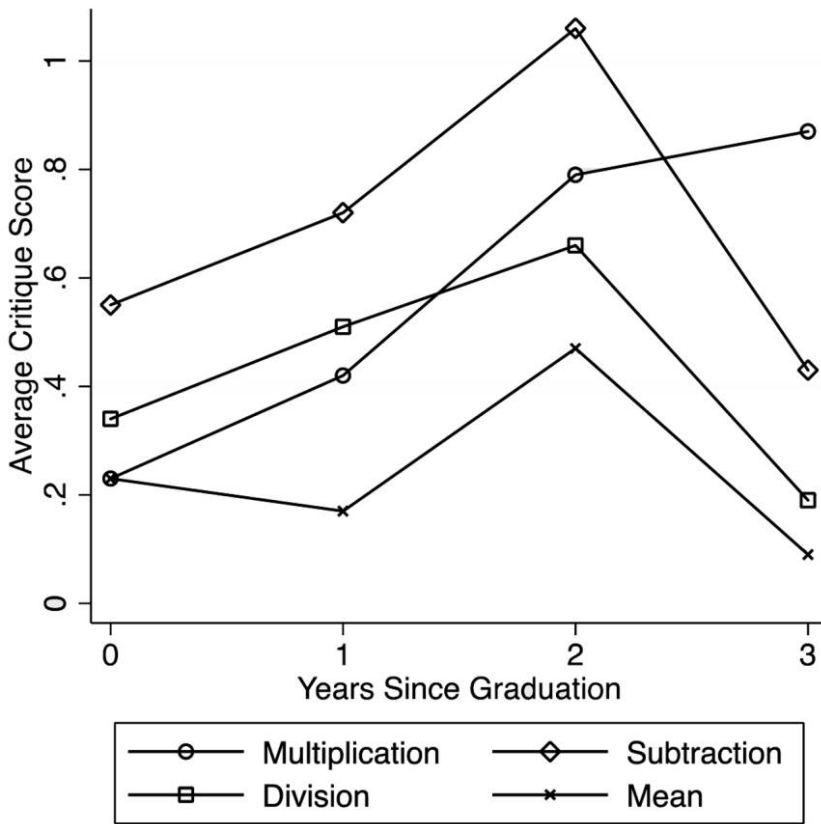


Figure 2. Changes in average critique scores for the four topics over time.

of-teaching task for these topics, we conducted a principal component analysis by specifying a single factor solution. We ran the analysis with the three target topics and then with all four topics for the two primary scores—the mathematical score and the critique score. For our hypothesis to be confirmed, the percentage of the variance explained by the three target topic model should exceed that explained by the four-topic model. Table 4 shows the results of this analysis for both the mathematical scores and the critique scores.

Each year, for both the mathematical and critique scores, the percentage of variance explained is higher when running the analysis for the three target topics than for all four topics. In other words, when adding the scores for the control topic, the percentage of explained variance in the performance on this task declines in all cases. Although some of the differences are not large, they are consistent across all comparisons. Our interpretation of these results is that the differences in performance can be accounted for by differences in the quality of participants' competence in the three target topics compared with the control topic. A direct and plausible explanation for these competence differences is the differential knowledge they acquired in the first two content courses in their preparation program. However, there are at least four alternative hypotheses that should be investigated before such a conclusion is viable.

Table 4. Percentage of Variance Explained Each Year Assuming a Single Underlying Factor for Three Target Topics versus All Four Topics

Score	Years since Graduation			
	0	1	2	3
Mathematics:				
Three target topics	57.4	66.5	62.2	58.7
All four topics	52.1	58.2	59.0	57.3
Critique:				
Three target topics	48.3	72.2	61.4	46.0
All four topics	36.8	63.3	54.6	38.5

Alternative Hypotheses

One of the challenges facing researchers who study the effects of teacher preparation on graduates' performance over time is the large number of potentially confounding factors between learning during the program (especially as freshmen) and eventual teaching competencies (Diez, 2010). Numerous factors could explain differences in performance on teaching-like tasks 4, 5, and 6 years later. This means it is unrealistic to *prove* causal links between teacher preparation and later performance. But it is possible to test some explanations for detected relationships other than teacher preparation. Four alternative hypotheses are as follows: (1) Opportunities to teach a topic might support higher performance on a related analysis-of-teaching task; (2) more professional development on a particular topic could support better analysis of teaching; (3) opportunities to teach from curricula that emphasize particular topics, including their conceptual underpinnings, could improve performance on teaching analysis tasks; and (4) the videos themselves might have created unequal opportunities to score well on the teaching analysis task.

Alternative Hypothesis 1: Teaching the topic. Perhaps the participants' higher scores on the target topics were not due to knowledge acquired during the preparation program but rather to the fact that more participants had opportunities to teach the target topics during the school year. Teaching a topic might yield knowledge that would help analyze someone else's teaching by looking for more conceptually based pedagogical interactions and more conceptual content. Table 5 shows the number of participants who taught each topic each year.

By inspection, it is clear there is no consistent relationship between the number of participants who taught a particular topic and the average video analysis score on that topic. In particular, the data do not support the hypothesis that participants scored

Table 5. Number of Participants ($N = 53$) Who Taught Each Topic Each Year since Graduation

Topic	Years since Graduation		
	1	2	3
Multiplication	27	32	29
Subtraction	26	27	26
Division	16	13	20
Mean	22	21	20

significantly higher on the target topics because more of them taught those topics. For example, although in no year did more participants teach dividing fractions than the other three topics, their scores on the mathematics and critique aspects of the dividing fractions task were not significantly lower than (sometimes significantly higher than) scores for the other three topics. Thus, this alternative hypothesis is not supported.

Alternative Hypothesis 2: Differences in professional development. Some teachers who enter the profession participate in professional development activities that aim to enhance their teaching of particular topics. These activities might have contributed to the performance differences in the analysis-of-teaching tasks. We cannot rule out this possibility because we do not have reliable data on the professional development activities in which the graduates participated during their first 3 years of teaching. However, as noted earlier, results on this task are part of a larger longitudinal project; as part of this project, we followed an earlier cohort of graduates into the field. Only 10% or fewer of the participants in this first cohort reported receiving professional development on any of the four mathematical topics studied here, and no particular topic was a more frequent focus of professional development (Morris & Hiebert, 2017). Because the first cohort and the cohort described in this study graduated only 1 year apart, we have no reason to believe the percentage of graduates in this study who received professional development on any of the topics was substantially higher or the distribution of professional development across topics was much different than reported by Cohort 1 participants.

Alternative Hypothesis 3: Teaching from a curriculum that emphasizes some topics. If a high concentration of graduates taught from a curriculum that provided better learning opportunities for themselves on the target topics in this study, the curriculum might explain their better performance on the target topic tasks. However, if, as a group, the graduates taught from a variety of curricula, this explanation is less likely. Although we do not have data on which curriculum each graduate who was teaching used each year, we do know they were spread among at least seven states and 21 school districts plus four charter schools and three private schools. In 2 of the 3 years, the distribution was even larger. This suggests the graduates were using a variety of curricula. Again, these data do not rule out the curriculum hypothesis, but they do reduce its likelihood.

Alternative Hypothesis 4: Different affordances provided by the video clips. To this point in the article, we have been treating the videotaped segments as essentially equivalent. Differences among topic scores have been interpreted as differences among graduates' competencies with these topics. However, this is an oversimplification. It is impossible to collect videos of actual classroom interactions that afford exactly the same opportunities to analyze them. For example, video clips that show more obvious instructional flaws might suggest more alternative instructional moves and, thus, yield higher critique scores.

To assess the potential effect of differences inherent in the video clips, we asked six expert mathematics educators to complete the same four tasks administered to our graduates. We defined *expert* as a person who has taught mathematics in elementary, middle, or high school and now is in a leadership position in school mathematics or is preparing elementary mathematics teachers. We reasoned that experts would have comparatively equivalent knowledge of all four topics, at least of the kind they would use to complete the teaching analysis tasks. In other words, if

the four video tasks provide similar opportunities to respond in ways captured by our rubrics, then our experts should score similarly on each of them. In contrast, differences in experts' scores across the four topics would signal different affordances of the task with respect to our rubrics.

Parenthetically, we did not expect the experts to reach ceiling effects on the teaching analysis tasks because the rubric was designed to capture particular aspects of the topics emphasized in the first two content courses of our preparation program. Our experts had not taken these courses. But we were interested only in whether the videos themselves, along with our scoring rubrics, created substantially different opportunities to score well for mathematics educators having similar knowledge of all four topics.

Although six respondents are too few to statistically compare performances across mathematics topics, it is clear that the mean scores of the experts' responses do not fit the same pattern as those of the program graduates. The experts' mean mathematics scores for multiplying two-digit whole numbers, subtracting fractions, dividing fractions, and finding the mean were 3.0, 2.3, 2.7, and 2.5, respectively. Their mean critique scores for the same topics were 2.3, 1.7, 1.8, and 1.7, respectively. Note that average scores on the finding-the-mean task are not substantially lower than scores for at least some of the target topics. This suggests the video clip for finding the mean is not more difficult to analyze, using our rubrics, than the clips for the three target topics.

Discussion

The findings of this longitudinal study suggest that specific mathematics content studied (even early) in a teacher preparation program can be accessed and used to complete a teaching-like task during graduates' first years of teaching. Graduates attended more to the mathematics content and to the conceptual aspects of the pedagogy if they had studied that content in the program, even 6 years earlier. The results seem to be explained better by knowledge acquired during the preparation program than by competing, alternative hypotheses.

In many ways, these data are exactly the kind of data policy makers would wish for if they wanted to show that teacher preparation can make a difference. Studying mathematics topics deeply is related to better performance on a teaching-like task on the same mathematics topics up to 6 years later. Graduates appear to retain and use the knowledge they acquired through their beginning years of teaching.

Our aim in this study was to investigate possible links between PSTs' content preparation and performance on a teaching-like task. As reviewed earlier, prior research has found stronger connections between pedagogy preparation and teaching skills than between content preparation and teaching skills. What might account for the connections reported here between studying content and better performance years later?

First, recall that the semester-long content courses develop relatively few topics—whole numbers and decimals in the first course, and fractions and proportional reasoning in the second course. This means these topics receive considerable attention and extensive development. The operations of multiplication, subtraction, and division (target topics for this study) are developed in multiple lessons across both

courses, first in the context of whole numbers and decimals and then in the context of fractions. We hypothesize that covering many mathematics topics, each for a brief time, is not sufficient for PSTs to retain and use the knowledge years later. We believe the considerable time spent developing the target topics was essential.

A second aspect of content preparation we hypothesize explains the findings is that the learning opportunities in both content courses are designed to support PSTs' development of MKT. In particular, PSTs have numerous opportunities to concretely and visually model operations on whole numbers, decimals, and fractions; to analyze children's solutions to whole number, decimal, and fraction problems; and to unpack the conceptual basis for standard and nonstandard algorithms for operating on whole numbers, decimals, and fractions. These are the kinds of competencies that would seem to be involved in analyzing teaching—the skill assessed in this study.

Finally, the courses are designed to emphasize and promote two pedagogical moves shown to support students' conceptual understanding (making key mathematical ideas explicit and providing students opportunities to grapple with these ideas). We conjecture that the participants in this study integrated their knowledge of these features of teaching with their knowledge of the content. As a result, we hypothesize that graduates left the program with well-developed MKT in the target topics and a clear sense of what is entailed in teaching the target topics for conceptual understanding, and these competencies supported their analyses of the video tasks on the target topics.

The answer to the question of whether knowledge developed by studying particular topics in the program can transfer to other topics not studied in the program is somewhat mixed. Task performances on the target topics were consistently higher than on the control topic and, in most cases, improved over time. But the pattern of improved performance over time on the target topics was mirrored in the control topic, at least on the mathematics scores. Perhaps the emphasis on conceptual understanding in the preparation program influenced graduates to focus on conceptual aspects of mathematics in their analysis of teaching, regardless of topic, but they simply had a head start on the topics developed in the program.

The relative absence of significant differences in performance just before graduation (Year 0) and the performance decline for some comparisons in Year 3 are not fully interpretable. With respect to the Year 0 results, it might be that participants became more attuned to (i.e., noticed) the important interactions in a classroom situation after they had experienced classroom interactions more intensively themselves. With respect to the decline in Year 3, perhaps when graduates were 6 years beyond when they studied the material in the program, they either forgot to use the knowledge on the tasks or the knowledge itself had deteriorated. Perhaps they did not use this kind of knowledge in their own teaching. We do not have information for all these participants about the nature of their teaching after graduation, so we cannot say whether regularly using the kind of knowledge acquired during their preparation program would keep it more active. But this seems like a reasonable conjecture.

Although the school subject we investigated in this study was mathematics, the key aspects of this study are not unique to mathematics. As long as the knowledge acquired about a subject can be specified at a level of detail to allow alignment be-

tween what PSTs are taught and the knowledge needed to analyze teaching, there would be no reason why similar investigations could not be conducted in other subject domains. In fact, the findings we report can inform hypotheses to be tested in other school subjects.

But many questions remain unanswered. First, are the relationships between teacher preparation and analysis-of-teaching skills specific for each topic? The results of this study suggest topic-specific relationships. If this is true, then teacher preparation really can matter, in the strictest sense. If PSTs do not develop particular subject matter knowledge for teaching, they are not likely to have such knowledge available to apply when they begin teaching.

A second, related question is whether a particular level or depth of knowledge must be acquired in a preparation program for it to be accessed and applied years later. We argued earlier that the depth with which the target topics were developed in the courses likely contributed to their performance years later. But is there a critical level of study? Could our graduates have spent less time on the target topics and performed as well? We cannot answer this knowledge-threshold question. The answer to this question is critical, however, because teacher educators must decide how much time to spend on which topics during a preparation program. This entails deciding not only which topics should be studied at what depth but also which topics should be omitted. These are difficult decisions, but we believe it is better for teacher educators to consider these questions deliberately rather than cover briefly all elementary school topics and hope graduates remember what they studied.

Finally, even though performance on a version of the teaching analysis task used here has been found to relate to high-quality teaching and, in turn, better student learning, data on classroom teaching and student learning would need to be collected to determine whether the pattern of performance reported here would be reflected in our graduates' teaching practices. Analyzing teaching, from both content and pedagogical perspectives, seems to be a critical teaching skill, but additional data are needed to determine whether studying content, even early in a preparation program, translates into more effective classroom teaching of that content after graduation.

Note

The research reported was supported in part by the National Science Foundation under grant DRL-0909661 and by the Institute of Education Science, U.S. Department of Education, through grant R305B100017. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the position, policy, or endorsement of the granting agencies. Thanks to Heather Gallivan for her assistance with data analysis. James Hiebert is Robert J. Barkley Professor at the School of Education, University of Delaware; Emily Miller is assistant professor at West Chester University; and Dawn Berk is assistant professor at the University of Delaware. Correspondence may be sent to James Hiebert at hiebert@udel.edu.

References

- Arbaugh, F., Ball, D. L., Grossman, P., Heller, D. E., & Monk, D. (2015). Dean's corner: Views on the state of teacher education in 2015. *Journal of Teacher Education*, *66*, 435–445.

- Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 83–104). Westport, CT: Ablex.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389–407.
- Berk, D., & Hiebert, J. (2009). Improving the mathematics preparation of elementary teachers, one lesson at a time. *Teachers and Teaching: Theory and Practice*, *15*, 337–356.
- Borko, H., & Putnam, R. T. (1996). Learning to teach. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 673–708). New York: Macmillan.
- Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, *31*, 416–440.
- Cochran-Smith, M., Villegas, A. M., Abrams, L., Chavez-Moreno, L., Mills, T., & Stern, R. (2015). Critiquing teacher preparation research: An overview of the field, Part II. *Journal of Teacher Education*, *66*, 109–121.
- Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Erlbaum.
- Diez, M. E. (2010). It is complicated: Unpacking the flow of teacher education's impact on student learning. *Journal of Teacher Education*, *61*, 441–450.
- Ericksen, F. (2011). On noticing teacher noticing. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 17–34). New York: Routledge.
- Feuer, M. J., Floden, R. E., Chudowsky, N., & Ahn, J. (2013). *Evaluation of teacher preparation programs: Purposes, methods, and policy options*. Washington, DC: National Academy of Education.
- Floden, R. E., & Meniketti, M. (2005). Research on the effects of coursework in the arts and sciences and in the foundations of education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 261–308). Mahwah, NJ: Erlbaum.
- Gallimore, R., Ermeling, B. A., Saunders, W. M., & Goldenberg, C. (2009). Moving the learning of teaching closer to practice: Teacher education implications of school-based inquiry teams. *Elementary School Journal*, *109*, 537–553.
- Gansle, K. A., Noell, G. H., & Burns, J. M. (2012). Do student achievement outcomes differ across teacher preparation programs? An analysis of teacher education in Louisiana. *Journal of Teacher Education*, *63*, 304–317.
- Henry, G. T., Campbell, S. L., Thompson, C. L., Patriarca, L. A., Luterbach, K. J., Lys, D. B., & Covington, V. M. (2013). The predictive validity of measures of teacher candidate programs and performance: Toward an evidence-based approach to teacher preparation. *Journal of Teacher Education*, *64*, 439–453.
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte, NC: Information Age.
- Hiebert, J., & Morris, A. K. (2009). Building a knowledge base for teacher education: An experience in K–8 mathematics teacher preparation. *Elementary School Journal*, *109*, 475–490.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*, 430–511.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371–406.
- Hill, H. C., Sleep, L., Lewis, J. M., & Ball, D. L. (2007). Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts? In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 111–155). Charlotte, NC: Information Age.
- Jacobs, V. R., Lamb, L. L. C., Philipp, R. A., & Schappelle, B. P. (2011). Deciding how to respond on the basis of children's understandings. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp

- (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 17–34). New York: Routledge.
- Kersting, N. (2008). Using video clips as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement*, **68**, 845–861.
- Kersting, N. B., Givvin, K., Sotelo, F., & Stigler, J. W. (2010). Teachers' analysis of classroom video predicts student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, **61**, 172–181.
- Kersting, N. B., Givvin, K. B., Thompson, B. K., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, **49**, 568–589.
- Levine, A. (2006, September). *Educating school teachers*. Washington, DC: Education Schools Project. Retrieved from http://www.edschools.org/teacher_report_release.htm
- Lincove, J. A., Osborne, C., Mills, N., & Bellows, L. (2015). Teacher preparation for profit or prestige: An analysis of a diverse market for teacher preparation. *Journal of Teacher Education*, **66**, 415–434.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics Review of Education*, **13**, 125–145.
- Morris, A. K., & Hiebert, J. (2017). Effects of teacher preparation courses: Do graduates use what they learned to plan mathematics lessons? *American Educational Research Journal*. Advance online publication. doi:10.3102/0002831217695217
- National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Academies.
- Ottmar, E. R., Rimm-Kaufman, S. E., Larsen, R. A., & Berry, R. Q. (2015). Mathematical knowledge for teaching, standards-based mathematics teaching practices, and student achievement in the context of the Responsive Classroom Approach. *American Educational Research Journal*, **52**, 787–821.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, **62**, 307–332.
- Santagata, R., & Angelici, G. (2010). Studying the impact of the lesson analysis framework on pre-service teachers' abilities to reflect on videos of classroom teaching. *Journal of Teacher Education*, **61**, 339–349.
- Santagata, R., Zannoni, C., & Stigler, J. (2007). The role of lesson analysis in preservice teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, **10**, 123–140.
- Schoenfeld, A. H. (2011). On noticing teacher noticing. In M. Sherin, V. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 223–237). New York: Routledge.
- Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (2011). Situating the study of teacher noticing. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 3–13). New York: Routledge.
- Sherin, M. G., & van Es, E. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education*, **13**, 475–491.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, **15**(2), 4–14.
- Star, J. R., & Strickland, S. K. (2008). Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education*, **11**, 107–125.
- Stocker, S. L. (2008). Using a video-based curriculum to develop a reflective stance in prospective mathematics teachers. *Journal of Mathematics Teacher Education*, **11**, 373–394.
- van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, **10**, 571–597.
- van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education*, **24**, 244–276.
- Zeichner, K., & Tabachnick, R. (1981). Are the effects of university teacher education "washed out" by school experience? *Journal of Teacher Education*, **32**, 7–11.