9-11-2014

Pre-Service Teachers Learning to Generate Evidence-Based Hypotheses on the Effects of Teaching on Student Learning

Cathery Yeh
Chapman University, yeh@chapman.edu

Rossella Santagata
University of California, Irvine

Recommended Citation

Yeh, C. & Santagata, R. (2014). Pre-service teachers learning to generate evidence-based hypotheses on the effects of teaching on student learning. *Journal of Teacher Education, 66*(1): 21-34. doi: 10.1177/0022487114549470

Pre-Service Teachers Learning to Generate Evidence-Based Hypotheses on the Effects of Teaching on Student Learning

Comments
This is a pre-copy-editing, author-produced PDF of an article accepted for publication in *Journal of Teacher Education*, volume 66, issue 1, in 2014 following peer review. The definitive publisher-authenticated version is available online at DOI: 10.1177/0022487114549470.

Journal of
Teacher Education

Pre-Service Teachers' Learning to Generate Evidence-Based Hypotheses about the Impact of Mathematics Teaching on Learning

Journal:	*Journal of Teacher Education*
Manuscript ID:	JTE-14-01-0047.R2
Manuscript Type:	Research/Empirical
Keywords:	Elementary Teacher Education, Mathematics Teacher Education, Preservice Teacher Education, Teacher Learning, Video Analysis
Additional keywords:	Learning to Teach, Analyzing Teaching

Abstract:	This study examines the development of a specific sub-skill for studying and improving teaching – the generation of hypotheses about the effects of teaching on student learning. Two groups of elementary pre-service teachers (PSTs) were compared: one group who attended a typical mathematics methods course and one who attended a course integrating analysis skills for learning from teaching. Data consists of PSTs' comments on video clips of mathematics instruction administered before and after course completion. Findings reveal that PSTs at the beginning of the program struggled to generate hypotheses with relevant evidence, often equating teacher behavior or student correct answers as evidence of student understanding. After course participation, PSTs who attended the course with integrated analysis skills significantly improved in their ability to generate hypotheses based on student evidence while their counterparts continued to display difficulties. Implications for teacher education and future research are considered.

Pre-Service Teachers' Learning to Generate Evidence-Based Hypotheses about the Impact of Mathematics Teaching on Learning

Introduction

Current education reforms in the U.S. press teachers for ambitious teaching practices that attend closely to all students' learning of important content (Lampert & Graziani, 2009; McDonald, Kazemi, & Kavanagh, 2013). This call is clear; what is less obvious is how teacher education programs can support pre-service teachers (PSTs) in the development of the necessary knowledge and skills.

This project contributes to our knowledge of experiences that might facilitate PSTs' learning of practices at the core of ambitious teaching. Typical models of teacher preparation attempt to impart a repertoire of expert knowledge and skills. However, the expectation of equipping PSTs with all the knowledge and skills necessary for ambitious teaching by graduation is unrealistic (Hammerness et al., 2005; Hiebert, Morris, Berk, & Jansen, 2007). Ambitious teaching is a complex endeavor and challenging for PSTs to learn, particularly given that most American teachers are graduates of the same education system we seek to improve. Therefore, we, like many in the field of teacher education, call for an alternative model of teacher preparation.

This model fosters many of the essential elements of ambitious teaching: PSTs learn to attend to student thinking, to develop student-focused practices, and to reason about strategies that assist students in reaching proficiency. The distinction in this alternative model lies in its conceptual basis, which stems from theories of how teachers learn (Brown, Bransford, Ferrara & Campione, 1983; Hawkins, 1973). PSTs are only at the *beginning* of their teaching/learning trajectories; most of their learning occurs once they enter the teaching profession. Therefore, in

conjunction with the development of skills for ambitious teaching, a central goal is to *learn to learn*.

Learning to learn from teaching is not a novel concept. Hawkins (1973) hypothesized, "It is possible to learn in two or three years the kind of practice which then leads to another twenty years of learning" (p.7). Most preparation programs engage in some form of reflective activity, usually in conjunction with PSTs' fieldwork experience. However, not all *reflection* on practice leads to *learning* from practice (Santagata & Guarino, 2011, Davis, 2006). A focus on *systematic* and *deliberate* analysis of practice may place PSTs on a different teaching and learning trajectory than that offered by traditional preparation models. At the center of this proposed trajectory are the development of instructional and reflective practices in which student learning drives decisions and sense making. Teachers should be deliberate in their plans to elicit and build on student learning during instruction and use evidence of student learning to reason about the effectiveness of their instructional decisions.

Historically, these types of practices have not been systematically taught during teacher preparation. Fieldwork experiences were often thought of as the settings in which PSTs could learn them. This study investigates the value added by systematic instruction targeting these practices. The larger project of which this study is part compares, through an experimental design, two mathematics methods courses: a course that follows a typical format centered on the development of mathematics content and pedagogy (i.e., the Math-Methods Course (MMC)) and another that integrates dispositions, knowledge, and skills for learning from practice into content and pedagogy (i.e., the Learning from Mathematics Teaching Methods Course (LMT)). All other experiences are kept constant: pre-service teachers attend the same courses and complete fieldwork. Field placements vary, but because PSTs are randomly assigned to them, variation

across both groups should be similar. PSTs' learning is investigated both short-term—during the teacher preparation program—and long-term—during their first three years of professional practice as classroom teachers. This study was conducted during the short-term phase and investigated PSTs' development of a specific set of skills needed in order to be able to learn from practice – the ability to generate evidence-based hypotheses about the impact of teaching on student learning.

Knowledge and Skills for Systematic Analysis of Teaching

Much of the conceptualization on dispositions, knowledge, and skills to learn from practice has been conducted by Hiebert and colleagues (Hiebert, Morris, & Glass, 2003; Hiebert, et al, 2007). They proposed four skills: (a) specifying lesson learning goals; (b) conducting empirical observations to collect evidence of student learning; (c) constructing hypotheses about the effects of teaching on students' learning; and (d) proposing improvements in teaching on the basis of these hypotheses. These skills must be integrated into teachers' daily planning, implementation, and reflection on their instruction and are supported by the edTPA performance-based assessment of teaching. Detailed information about the edTPA can be found at http://edtpa.aacte.org/.

Although the conceptualization is well-developed, empirical studies remain "too scarce to confirm that PSTs can acquire the skills necessary to analyze teaching" (Hiebert, et al., 2007, p. 58). Only a few studies have reported PSTs' learning of the specific skills necessary to learn from teaching. Although this study focuses on skill 3 in the framework – the ability to construct hypotheses about the effects of teaching on students' learning,—in the subsequent sections, we will describe all skills and summarize study findings that examined their development. Skill 3 in fact builds on skill 1 and 2 and directly impacts skill 4.

Skill 1: Specifying the lesson learning goals. The ability to construct evidence-based hypotheses begins with descriptions of the mathematical learning goals (Hiebert et al., 2007; Morris, Hiebert, & Spitzer, 2009). Unless teachers are clear about what they intend students to learn, it is difficult to begin to plan or examine the impact of instruction on learning.

Take for example a lesson on subtraction. A teacher could phrase the learning goal as: students will be able to solve subtraction word problems. The teacher could further specify the goal: students will be able to interpret subtraction as both the removal of a smaller quantity from a larger quantity and as a comparison of a smaller quantity to a larger quantity (Berk & Hiebert, 2009). These two goals will lead to different assessments of student learning. Being clear about learning goals means "unpacking" a topic into its key ideas. Although the knowledge and skills needed to unpack learning goals is largely mathematical, Morris and colleagues (2009) found that content knowledge alone did not always translate into the ability to identify a lesson's key mathematical ideas. Rather, interventions targeting specifically this ability are necessary.

Skill 2: Conducting empirical observations to collect evidence of student learning. After specifying the learning goals, teachers need to collect evidence of student learning. This process involves: (a) recognizing that only evidence of students' learning can be used to justify the lesson effectiveness; (b) distinguishing between relevant and irrelevant evidence; and (c) knowing how to identify moments where evidence could be collected (Hiebert et al., 2007).

Previous studies suggest that PSTs enter education programs with misconceptions about what constitutes evidence of learning. PSTs often assess teaching effectiveness based on teacher strategies rather than on student responses (Hiebert & Stigler, 2000; Morris, 2006; Santagata, Zannoni, & Stigler, 2007). Even when the attention centers on students, over-attribution of conceptual understanding often occurs as PSTs equate correct answers or the completion of a set

of procedural steps as evidence of conceptual understanding. (Bartell, Webel, Bowen, & Dyson; 2012; Spitzer et al., 2010). For example, Morris (2006) found that PSTs were willing to accept a student's correct use of the area formula of a triangle as evidence of conceptual understanding rather than evidence of procedural knowledge.

Research suggests that it may be possible to help PSTs to attend appropriately to evidence of student learning under certain conditions. Among others (Jacobs & Philipp, 2004; Jansen & Spitzer, 2009; Morris, 2006), Spitzer and colleagues (2010) investigated the effects of an intervention aimed at addressing this issue. Bartell and colleagues (2012) also found that PSTs improved in distinguishing between evidence of procedural skill and conceptual understanding after an intervention centered on the analysis of student thinking.

Skill 3: Constructing hypotheses about the effects of teaching on students' learning. Formulating hypotheses about what aspects of instruction led to changes in student learning serves as a critical component of the analysis process. The generation of hypotheses shifts the lesson from being a learning experience solely for the students to being a learning experience for the teacher as well (Hiebert et al., 2003).

Skill three builds on skills one and two. Hypotheses are more likely to lead to improvements in student learning when they are (a) made about students' progress toward the learning goals; and (b) based on relevant and revealing evidence of students' learning and teacher's instructional decision-making (Hiebert et al., 2007; Morris, et al., 2009). Therefore, PSTs struggling with skill one and/or two would not be able to form strong hypotheses.

Prior studies show that teachers have a tendency to assume that a lesson is effective and to fill in the "holes" in evidence using principles of effective teaching, instead of actively

collecting evidence of student learning (Fernandez, Cannon, & Chokshi, 2003; Hiebert & Stigler, 2000; Santagata et al., 2007).

Very little is known about experiences that might help teachers identify and interpret instructional effectiveness based on evidence of student learning. The few existing studies have examined the development of the ability to integrate different elements of teaching when making sense of instruction. For example, van Es and Sherin (2002) engaged PSTs in a three month video-based intervention that required attending closely to episodes of classroom teaching and providing interpretations. PSTs shifted from primarily describing and evaluating practices to producing interpretive analyses that integrated student thinking, teacher's role, and classroom discourse and used these analyses to inform their evaluation of teaching effectiveness. Davis (2006) found similar results. She designed a series of reflective on-line, journal assignments in her undergraduate science methods course. Students wrote weekly about their teaching experience and received feedback specifically targeting the relations among student learning, subject matter knowledge, assessment, and instruction. An analysis of PSTs' self-reflections revealed that over time these became more analytical and integrated.

Stockero (2008) and Moore-Russo and Wilsey (2014) used representations of teaching to promote reflection in their mathematics methods courses. Stockero (2008) utilized a video-case curriculum, while Moore-Russo and Wilsey (2014) utilized animations. Both interventions introduced minimal prompts to guide PST viewing. Stockero (2008) found that PSTs learned to consider multiple interpretations of student thinking and to develop a more tentative stance in their inquiry. Moore-Russo and Wilsey (2014) found PSTs varied in their ability to recognize the complexity of teaching and learning and the interplay between teacher, students, subject

matter, and assessment. Their findings suggest that some scaffolding may be necessary to support analysis skills development.

Lastly, in a study by Santagata and Angelici (2010), PSTs significantly improved their ability to analyze instruction after a learning experience with an analytic framework centered on integrating elements of instruction in comparison to a group that used a framework focused on the evaluation of separate elements of instruction. Participants using the integrated framework were more likely to link lesson evaluations to concrete evidence, to focus on students' thinking and learning, and to provide suggestions for improvement.

Skill 4: Proposing improvements in teaching on the basis of the hypotheses. The hypotheses about the effects of teaching on student learning form the basis for making evidence-based decisions for improving instruction. In other words, the formulation of instructional revisions requires following the implicit or explicit recommendations contained in the hypotheses. When implemented, instructional improvements conclude the cycle of reflection by connecting analysis back to instructional practice.

The quality of the proposed improvements depends on whether the revised instruction facilitated students' progress toward the learning goals (Hiebert et al., 2007); this requires testing revisions in real classrooms. This can be difficult to apply in most preparation programs as PSTs have limited student access. Yet, it is possible for PSTs to practice proposing suggestions and providing rationales for improvement.

At different points in the analysis process delineated above, teachers have opportunities to learn from practice. For example, they may learn about student thinking and understanding of mathematical ideas; instructional strategies that are helpful in teaching particular mathematical

content and ones particularly helpful in making student thinking visible; and finally, the need to rethink a strategy used because it did not lead to the desired learning outcomes.

Study Aims and Research Questions

The reviewed studies provide some evidence that the skills for systematic analysis of practice can be taught. However, the majority of the existing literature focuses on the development of PSTs' ability to analyze student thinking (Bartell, et al., 2012; Morris, 2006; Spitzer et al., 2010), while only a few focus on the ability to generate hypotheses about the effectiveness of teaching. This skill, we would argue, is necessary if teachers want to learn from their current practice to inform future practices.

Thus the study goals are to: (a) better understand PSTs' ability to generate evidence-based hypotheses, and (b) investigate ways this ability can be improved during teacher preparation. By delineating a learning trajectory, findings will contribute to the development of a theory of teacher learning from practice that specifies typical initial understandings, its evolvement over time, and what learning experiences might facilitate this process. The study addresses the following questions:

1) What is the quality of PSTs' beginning ability to generate evidence-based hypotheses about the effects of instruction on student learning? What type of reasoning do PSTs display?
2) Can math-methods course experiences improve PSTs' ability to generate evidence-based hypotheses? And, if so, does a course that integrates instruction on systematic analysis of teaching improve PSTs' ability to a greater extent than a more typical course targeting content and pedagogy?

Study Design and Methods

Study Participants

Study participants attended a fifth-year post baccalaureate elementary teacher preparation program at a public West Coast University. Seventy-seven PSTs were randomly assigned to one of the two math methods courses while they attended the same courses for all other disciplines taught in the program. Since mentor teachers and fieldwork settings may vary in the extent to which they expose and apprentice PSTs in reflecting practices, PSTs were randomly assigned to fieldwork placements. This allowed us to investigate the value added by systematic instruction on analysis of teaching independently from variation in fieldwork experiences.

Study recruitment for voluntary participation occurred during program orientation. Of the 77 enrolled in the program, 60 consented to participate and completed the pretest and posttest measures. Twenty-nine attended the LMT course, and 31 attended the MMC course. There were no significant differences between the two groups in undergraduate degree and prior teaching experience. The majority held a baccalaureate degree as their highest level of education and had minimal teaching experience. Of the 60 participants, 53 were female and 7 male. The ethnic composition of the participants was 47% Caucasian, 43% Asian, 7% Hispanic/Latino, and 3% other.

The Mathematics Methods Course

The MMC was designed to support PSTs' development of content knowledge and pedagogical skills for all mathematics domains in K-6 teaching. In the two-quarter course, PSTs received approximately 50 hours of instruction structured around the mathematics methods textbook authored by Van de Walle, Karp, and Bay-Williams and published by Pearson Education (2008), which is utilized by 70% of U.S. mathematics methods courses (K. Scheyving, personal communication, April 15, 2010). The course combined opportunities to develop an

understanding of children's mathematics learning, problem-based instruction, and instructional planning and assessment. The MMC fostered a "learn-by-doing" approach; PSTs wrestled with mathematical concepts, explored physical models, and practiced explaining their reasoning to their colleagues. A few reflective activities (i.e. reflections on the implementation of math lessons during fieldwork) were infrequent and not guided by a systematic framework.

The Learning from Mathematics Teaching Methods Course

The LMT course also focused on the development of content and pedagogy and utilized the same textbook. This course concentrated on number sense and fractions and on skills for learning from teaching using a framework, the Lesson Analysis Framework, developed by Santagata et al. (2007) to provide PSTs opportunities to practice the four analysis skills outlined above. This framework consists of a series of questions that guide teachers' analysis: What is the main learning goal? Are the students making progress towards the learning goal? What evidence do you have that students are making progress? What evidence is missing? Which instructional strategies supported students' progress towards the goal and which did not? What alternative strategies would better assist students' progress towards the goal?

The LMT curriculum combined analysis-of-teaching activities and opportunities to practice important instructional routines (see Table 1 for a summary of key tasks and instructional resources). Video was used extensively to provide images of ambitious teaching and to facilitate and share the analysis process. PSTs reviewed videotapes of individual student interviews and teaching episodes and, through collaborative analysis guided by the Lesson Analysis Framework, highlighted student thinking and the interrelation between teachers' decisions and student learning. PSTs were also provided opportunities to plan and enact instructional activities that make student thinking visible and to analyze their enactments

collaboratively. Instructional tasks around the Lesson Analysis Framework were planned to gradually scaffold PSTs from supported to independent analyses and from analyses of others to analysis of their own teaching. The end goal was for this systematic process of analysis to become a repertoire of daily practice.

Insert Table 1 here.

Measure: The Classroom Video Analysis Assessment

Participants completed the Classroom Video Analysis (CVA), an assessment developed to capture teacher knowledge for teaching mathematics (Kersting, 2008), prior to and at completion of the methods courses. The CVA consists of a series of brief video clips (one to three minutes in length) that portray either teacher assistance during student independent work or student mistakes that ensue discussion during whole-class instruction. Teachers view the clips via an interactive, Web-based platform and respond to the prompt, "discuss how the teacher and the student(s) interact around the mathematical content." This measure was chosen because it has been validated in previous studies that found the quality of teachers' CVA written comments predicted student learning through quality of instruction (Kersting, Givvin, Thompson, Santagata, & Stigler, 2012).

Four series of scoring rubrics are applied to PSTs' written responses. PSTs receive a global score capturing their overall ability to analyze teaching and a score on each sub-scale. The scoring rubrics assess teachers' abilities to: (1) attend to the mathematical content in the clip; (2) attend to student thinking; (3) elaborate on what they see beyond pure descriptions; and (4) propose instructional improvements. Tested video clips are not publicly released, but similar clips with teacher comments and scoring examples are accessible at: http://www.teknoclips.org/examples/.

As part of a larger project, PSTs viewed and commented on 10 clips, and their responses were scored according to the CVA rubrics. For this study, we examined commentaries to four clips and developed a new set of codes to capture PSTs' ability to generate evidence-based hypotheses. The four clips portray instruction on the following topics: (1) double-digit multiplication, (2) regrouping in subtraction, (3) comparing fractions with a number line, and (4) equivalent fractions. As we were interested in capturing PSTs' abilities to generate hypotheses, we selected clips that contained a clear student-teacher interaction that allowed a window into student thinking.

Procedures

New codes were developed through a two-phase process. We first distinguished between responses that included valid reasoning about the effects of teaching on learning and those that did not. Then we categorized both kinds of responses into types to investigate more deeply both productive reasoning and common challenges. Figure 1 illustrates the coding process. Following is a description of each coding phase.

Insert Figure 1 here.

Phase I: Measuring pre-service teacher ability to generate hypotheses. Hypotheses are tentative claims about how an instructional event influenced the intended learning goal (Hiebert, 2003, 2007). To investigate PSTs' ability to generate hypotheses, we thus looked in each response for the presence or absence of a justified claim. Drawing from existing literature on PSTs' analysis skills (Bartell et al., 2012; Hiebert et al., 2007; Morris, 2006; Santagata et al, 2007; Santagata & van Es, 2010), we considered justified claims only responses that: (a) provided an analysis about the impact of the teacher's actions on student learning of the mathematics content; (b) depicted the mathematics correctly; and (c) cited relevant evidence of

student learning to justify the claim or discussed in detail the learning opportunities provided by particular instructional moves. These three components were deemed necessary to produce the kind of reasoning about the impact of teaching on learning that is conducive to a productive reflection on teaching. To be clear, if a response included some reasoning about the effects of teaching on learning but did not reference any student-based evidence, it was not considered a justified claim. Similarly, if a response described the mathematics in the lesson inaccurately (e.g. praising the teacher's emphasis on place value when the teacher inaccurately called numbers by its digit value instead of its place value), it was not considered as a justified claim because it led to incorrect reasoning about the effects of teaching on learning and sometimes to incorrect teaching suggestions.

Responses that included justified claims were given a score of 1, and responses that did not were given a score of 0. For each participant, four responses (i.e., a comment to each of the four clips) were scored, thus participants' scores ranged from 0 to 4. Repeated-measures analysis of variance was conducted to study the impact of the methods courses on PSTs' ability to generate claims. One-way ANOVAs were conducted to test for significant group differences at pre- and post-test and paired *t*-tests to assess for changes over time within each group.

Phase II: Categorizing response types. We were interested in two questions during the second coding phase: What types of claims do PSTs generate? What types of evidence do they use to support their analysis of the effects of instruction?

We looked separately at responses that generated a justified claim and responses that did not. Categories were developed by reviewing and separating all responses into qualitatively different types. Examining PSTs' initial reasoning, particularly when a valid claim is not produced, is important. It allows us to characterize their initial understandings, the first step in

their learning trajectory. In addition, it facilitates the design of experiences that build on and develop PSTs' analysis skills.

Types of reasoning when a justified claim was absent. Responses that obtained a score of 0 during the first coding phase fell into three categories: (a) descriptive, (b) missing relevant evidence, or (c) inaccurate depiction of the mathematics (see Table 2). Descriptive responses read like summaries and/or included comments on only certain distinct aspects of teaching and learning without considering their interconnectedness. These responses often discussed some teacher action and certain student behavior but did not reference the mathematics at the core of the clip.

Responses that did not include relevant evidence were of two types. The first included reasoning about the effectiveness of teacher decisions but without student-based evidence or justification (e.g. the teacher did a good job asking questions to help students understand). The second included evidence to justify claims of effectiveness, but this was irrelevant to the lesson learning goal (e.g. equating students' procedural fluency as evidence of conceptual understanding) or not revealing of students' learning (e.g. taking students' nods as evidence of understanding).

Finally, sometimes PSTs misinterpreted the interaction portrayed in the clip by depicting inaccurately the mathematics (e.g., the statement "all fractions are smaller than 1" would be categorized as a mathematical error). Likewise, a PST who endorsed with her response a mathematical idea not evident in the clip (e.g., a response described how the teacher directly addressed place value concepts when this was not seen in the clip) was also included in this category. Responses with inaccurate depiction of the mathematics were sometimes also descriptive or included irrelevant evidence. In these instances, they were coded as "inaccurate

depiction of the mathematics" because we considered understanding of the mathematics portrayed in the clip as the first step that guides all subsequent reasoning in the analysis process. The table below summarizes these response categories.

Insert Table 2 here.

Types of claims. Responses that produced a justified claim, thus, obtaining a score of 1 during the first coding phase, were considered together and coded according to the type of reasoning they included. Three categories captured the differences that were observed in the data: (a) principles of effective teaching (b) student learning; and (c) request for more evidence (see Table 3).

Claim Type: Principles of effective teaching or evidence of student learning**.** All the responses that included a justified claim about the effects of instruction on students' learning discussed all three aspects of teaching and learning: the teacher, the student, and the mathematics content. However, some PSTs justified their claims by drawing on principles of effective mathematics instruction while others on evidence of student learning. Student-focused claims began by analyzing student learning of the mathematics at a particular instance of the clip and discussed how specific instructional moves seemed to support learning. Claims based on principles of effective teaching began with an analysis of the teacher's moves and discussed the opportunities for mathematics learning that particular instructional decisions afforded. Here, PSTs did not attend to evidence of student learning directly visible in the video; rather they described the potential learning outcomes that might result from particular teacher decisions (see Table 3 for examples).

Request for more evidence. A few of the responses included an effort to reason about the effects of teaching on learning but lamented the fact that the evidence present in the clip was

insufficient to generate a valid claim. These responses usually began with an analysis of the student action or discussed specific teacher moves but directly stated that there was insufficient evidence present in the clip to make a claim about student thinking or learning. A third category was developed to capture these responses. Scores and coding categories described above were applied to both pre- and post-analyses.

Insert Table 3 here.

Inter-Rater Reliability

Both authors, blind to group membership, worked together to develop the coding system. This was done by analyzing 10% of responses, randomly chosen, until coding consensus was reached. Each researcher then coded a new randomly-selected 10% of responses. Inter-coder reliability was checked and consensus reached on disagreements. Finally, another randomly-selected 10% of responses was coded and reliability checked once more. Reliability, expressed in terms of % agreement, reached 80% during the first coding step and 85% during the second for both phase I and II codes. At this point, the first author coded the remaining pre- and post-test responses, but every time she encountered a response that was difficult to score, this was discussed with the second author until consensus was reached.

Results

Findings are presented in two sections. The first section examines PSTs' abilities to generate a justified claim at the beginning of the preparation program. The second section focuses on PSTs' abilities at course completion and compares participants who attended the two different courses.

The average comment length was 70 words. Sample commentaries are included to illustrate patterns found in the data analysis and to highlight differences and similarities between response

types. All selected commentaries respond to the subtraction clip that focused on developing understanding of regrouping, that is the exchanging of 1 in a place-value position for 10 ones in the position to the right. An understanding of regrouping requires students to flexibly see "a ten" as both a single entity and as a set of ten units.

PSTs' Ability to Generate Claims at the Beginning of the Program

PSTs' video commentaries for each of the four clips were first scored a 0 or 1 based on the presence or absence of a justified claim; thus participant scores ranged from 0 to 4. The average score at the beginning of the program was 1.22 ($M = 0.30$ per clip), indicating that PSTs in general failed to produce a justified claim. A two-sample t-test of PSTs' pretest scores found no significant difference between the MMC and the LMT group (MMC pretest $M = 1.26$, $SD = 1.06$, LMT pretest $M = 1.17$, $SD = 0.93$), $t(58) = 1.49$, $p = 0.23$), therefore pre-test findings for the two groups will be discussed as combined (for percentages by group, see Tables 4 and 5). The second coding phase examined the type of reasoning PSTs included in their responses.

Categorizing Pre-service Teacher Responses at Pre-Test

Types of reasoning about teaching and learning. Seventy percent of all pretest responses did not generate a justified claim. Of these, 37.3 % were coded as "descriptive." A sample commentary is from Ryan, a LMT group teacher candidate:

> The teacher helps the student draw in the bars, lines and circles and asks the student questions. Another student is there with the physical blocks, bars, and sticks but the focus is on the girl solving the problem.

Ryan's response consisted largely of a descriptive account of the teacher action and a brief reference to a student. The student was only mentioned at the end of the response where he wrote "the focus is on the girl solving the problem." There was no mention of any particular student idea. As opposed to a claim that elaborates on the impact of a teacher's action on student

learning, Ryan's response did not contain an integration nor an analysis of the cause-effect relationship between teaching and learning (i.e., the teacher, the student, and the mathematics).

Fifty-one percent of responses that did not generate a justified claim utilized irrelevant evidence to support their analyses. An example is Sandra's response:

> This is a really good strategy for the children when working on subtraction problems for the first time. The girl is learning about borrowing and carrying over numbers when you need to subtract. By using both writing visually and 3-D board pieces, a child understands how math works and can use it for real-life experiences.

At first glimpse, Sandra's response seems to be focused on the *student*. A careful review reveals that Sandra produced a claim on the impact of the *teacher's* choices—using visuals and manipulatives – on a student's learning. However, the claim is unjustified. Using visual representations and the base-ten blocks *can* support a better connection of the written subtraction algorithm from the visual representation. Nonetheless, the use of visuals alone does not assure understanding. The effectiveness of the tool in supporting student learning can only be assessed based on evidence from the student, but Sandra does not provide any evidence to support her claim.

Eleven percent of responses categorized as not generating a justified claim inaccurately depicted the mathematics in the clip. Responses that fell in this category provided an inaccurate description of the learning goal or endorsed a teacher action that was mathematically incorrect in the clip. The number and percentages of each type of responses that did not include a justified claim are summarized in Table 4.

Although the majority of pretest responses remained at a low level of sophistication, a few responses produced claims about the effects of teaching on student learning.

Insert Table 4 here.

Types of claims. Thirty percent of all pretest responses provided justified claims. Of these, 33.8% of responses generated a claim focused on student learning. These responses analyzed evidence of student(s)' learning (i.e., specific student action, talk, or written work) as a result of a teacher's instructional move. Approximately 50% of responses generated claims about the potential impact of an instructional decision on student learning based on principles of effective teaching. One such example is a response from Jerome in the LMT group:

> I thought the teacher guided the student in solving the problem. She didn't necessarily give the student the answer. She asked the student to model the problem and then had [her] write the traditional algorithm. I think this helps the student to develop a procedural understanding as well as a conceptual understanding.

Jerome generated a clear claim about the opportunities for student learning afforded by the teacher's decision to have the student directly model the meaning of the operation prior to writing the pencil-and-paper algorithm. Jerome and Sandra's response are similar in that they both produce a claim about the impact of the teacher's instructional decision-making on student learning but vary in the degree of certainty of the impact itself. Sandra stated that "*a child understands"* when visual referents are used to teach the traditional algorithm while Jerome wrote "*I think this helps the student to develop a procedural understanding*." Without direct evidence of learning from the student, the impact of these instructional features must be viewed as potential opportunities for learning rather than evidence of learning.

Finally, 13% of responses in this category discuss the need for more evidence of student learning to make claims about the impact of teaching. The number and percentage of each type of response within those that generated a justified claim are summarized in Table 5.

Insert Table 5 here.

Effects of the Methods Courses

Repeated-measures analysis of variance revealed a group x time interaction effect ($F(1,58) = 22.674$, $p < .001$, $\eta^2 = 0.37$) (See Figure 2). While, as mentioned above, groups did not differ at pretest (MMC pretest $M = 1.26$, $SD = 1.06$, LMT pretest $M = 1.17$, SD = 0.928, $t(58) = 1.48$, $p = 0.23$), the LMT group significantly outperformed the MMC group at posttest (MMC posttest $M = 1.48$, $SD = 1.09$, LMT posttest $M = 3.14$, $SD = 0.58$, $F(1, 58) = 16.683$, $p < .001$, $\eta^2 = 0.223$).

A paired *t*-test revealed that the LMT group performed significantly better on the posttest than on the pretest (LMT pretest $M = 1.17$, $SD = 0.928$, LMT posttest $M = 3.14$, $SD = 0.581$, $t(28) = -10.407$, p < .001). This indicates that the LMT participants' responses shifted from general descriptions, claims without relevant evidence, or inaccurate depictions of the clip (LMT pretest $M = .29$ per clip) to include justified claims (LMT posttest $M = 0.78$ per clip). Whereas, the quality of the MMC group comments did not improve significantly over time (MMC pretest $M = 1.26$, $SD = 1.06$, MMC posttest $M = 1.48$, $SD = 1.09$, $t(30) = -0.98$, $p = 0.33$). On average, the MMC group failed to generate justified claims for half of the responses on both the pre- and post-tests (i.e., MMC pretest M= 0.32 per clip, MMC posttest M = 0.37 per clip).

Insert Figure 2 here

Types of claims generated at posttest by each group. Table 6 and 7 display the types of claims generated by the two groups at posttest. The following sections discuss the most prevalent response types for each group.

MMC response at post-test. MMC participants displayed at posttest the same types of reasoning displayed at pretest. The majority of the MMC responses did not generate a justified claim. The types of responses also did not change (see Table 6). This same pattern is visible in the responses that included justified claims (see Table 7).

Sandra's posttest response exemplifies this consistency in reasoning from pre- to post-test. As described above, Sandra's pre-test response produced a claim on the effects of teaching on student learning but did not justify the claim using relevant evidence. Sandra wrote in her posttest response to the same clip:

> I really like the teacher-made worksheet that really helps students make sense of the numbers there. [Since] students see that there are more pieces in the ones place to take away, students learn to borrow numbers from the tens place. I definitely like the use of concrete material as well (using block manipulatives). The teacher is great at helping the student guide her thinking and comment on what she is doing, as well as guide her thinking. The teacher is patient and waits to see what she gets.

Sandra's posttest response displayed a deeper analysis of the mathematics in the clip than on her pretest. She mentioned the importance of direct modeling to support the understanding of regrouping in the traditional subtraction algorithm. Although a richer description of the mathematics was provided, Sandra struggled again to recognize evidence *relevant* to student learning. Her claim is based on the teacher instead of the student. Just as in the pretest, her response displayed little skepticism about the lesson's effectiveness and provided no analysis of the student responses visible in the clip.

LMT response at post-test**.** LMT participants displayed more sophisticated reasoning at posttest. The majority of LMT participants' responses generated a justified claim. Within the category of unjustified claims, there was a decrease in frequency for all three response types from pre- to post-test (see Table 6). Whilst in the category of justified claims, there was an increase in frequency for all response types (see Table 7). The majority of LMT responses displayed an analysis similar to the one illustrated in Jerome's pretest analysis. They demonstrated an explicit concern for student learning but did not provide an assessment of student understanding. Instead, these claims focused on particular teaching moves and their affordance for learning.

A substantial number of LMT responses generated claims focused on student thinking/learning. Thirty-two of the LMT posttest responses as opposed to 11 pretest responses produced this type of claim. An example is from Meredith's posttest response:

> It's a good idea for the student to see borrowing with a visual representation, but it seems that the teacher is simply walking her through the steps of the algorithm. I don't feel that the student would have much recall of what is happening. The student understands place value and that you can split up numbers like 138 into hundreds, tens, and ones but she doesn't seem to understand that tens can be split up into ones, hundreds into tens, etc. In this case, I think physical manipulatives would be more valuable than drawings and I think the student would benefit greatly from working with a partner who understands that tens can be split up into ones.

Meredith's response displayed skepticism in the effectiveness of instruction on learning. She justified her skepticism with an analysis of the student's learning as she highlighted the use of the place-value mat. Her response attended carefully to student thinking and demonstrated her ability to focus on the student instead of the teacher.

Request for more evidence**.** The LMT posttest responses categorized as requesting more evidence doubled from the pretest and in comparison to the MMC posttest responses (see Table 7). A representative response is from Sam, an LMT-group PST:

> I really liked this worksheet. It allowed students to work on the problem both in number form and [provides] a visual representation with the base ten blocks. The student did not contribute much to the conversation, and it seemed like the teacher was doing most of the talking. Because the student did not talk much, it was unclear whether she was understanding what she was doing when she was "borrowing" in the subtraction problem.

Sam's response demonstrated a critical stance towards teaching as well as the evidence available in the clip. Sam began his response by mentioning the importance of using visuals to support understanding of regrouping in subtraction. He then critiqued the available evidence. He described the student's limited response and indicated that there was not enough evidence to say whether the student understood. His response demonstrated a recognition that the use of conceptual referents does not equate to conceptual understanding. In this case, the evidence

provided in the clip was not sufficient for him to produce a strong claim about student understanding.

Discussion

It has become a goal of many teacher education programs to help teachers become more deliberate in their practices. A growing body of literature suggests that PSTs can learn to attend to evidence of student thinking (Bartell, et al., 2012; Morris, 2006; Spitzer et al., 2010). The ability to reason about the relation between teaching and learning has been less investigated. The few studies that have considered the generation of evidence-based hypotheses lay important groundwork for considering the affordances of various representations of teaching (Davis, 2006; Moore-Russo & Wilsey, 2014; Stockero, 2008; van Es & Sherin, 2002); however, with the exception of the study by Santagata and Angelici (2010), these studies have measured changes in PSTs' analysis skills without investigating whether more typical instruction may lead to similar results.

What's left to wonder is if PSTs develop these skills naturally from university courses and through fieldwork interactions. This study, by keeping constant all courses PSTs attended with the exception of the math methods courses and by randomizing fieldwork placements, was designed to examine this question. We begin with a discussion of the pretest results.

The Nature of Pre-service Teacher Analysis Abilities at the Onset of Teacher Preparation

Just as teachers are better equipped to promote student learning when they understand student's competencies, understanding PSTs' beginning analysis skills may inform the design of experiences that promote PSTs' further development. A series of implications for teacher educators derive from these findings.

First, a number of PSTs, as exemplified by Jerome's response, was able at program beginning to attend to both teachers and students and to make claims about the teaching-learning relations. However, these responses were inconsistent and scarce. The majority of pretest responses displayed misconceptions consistent with findings from prior studies (Davis, 2006; Morris, 2006; Spitzer et al., 2010) and provides additional evidence on this under-studied skill. Most PSTs displayed responses consisting of a clip summary or made claims using teacher-based evidence. Sandra's response is an example of such an analysis and reflects the assumption that "students learn what the teacher explains." This is problematic, as a teacher that assumes students learn what is said is likely to direct attention primarily toward her/his action instead of attending to the details of student thinking. Even when PSTs attended to the student, their responses revealed difficulties in identifying evidence that was both *revealing* of student learning and *relevant* to the lesson goal. Often, PSTs used evidence of students' procedural fluency as evidence of conceptual understanding. These findings suggest that PSTs need structured opportunities to analyze both examples and non-examples of student-based evidence and discussion of differences between evidence of procedural knowledge and of conceptual understanding.

Effect of the Methods Courses

Given that PSTs enter preparation programs with difficulties generating evidenced-based claims, can these skills be improved through teacher preparation and what experiences are necessary to do so? The MMC group showed an increase in the number of evidence-based hypotheses produced at posttest but misconceptions related to evidence-based reasoning about instruction continued to be an issue. Thirty-six of 124 MMC posttest responses struggled to recognize what would constitute relevant and revealing evidence of student learning. As

exemplified by Sandra's response, the MMC participants continued to justify lesson effectiveness using irrelevant evidence, citing teachers' strategies or the use of conceptual features of instruction. Similar to study findings by Moore-Russo and Wilsey (2014), there were variances among individual participants in hypothesis quality. As our participants were randomly assigned to their field placements, these variations might be due to differences in fieldwork settings with some better supporting student-centered instruction and reflection on teaching. On average, the MMC group did not show significant gains from pre- to post-test; thus this suggests that field experiences alone may not be sufficient to develop the kinds of analysis skills discussed in this paper.

Second, our findings provide support for teacher preparation experiences like those included in the LMT course and other studies (Santagata & Angelici, 2010; Davis, 2006) that combine instruction on systematic analysis and pedagogy. Study findings are promising as 75% of LMT posttest responses included justified claims. The responses from Meredith and Sam exemplify this type of analysis.

At the same time, these findings suggest areas of improvement. First, although the LMT course was successful in moving PSTs' away from descriptive and unsupported claims, at posttest the majority of the LMT PSTs generated claims, like Jerome's response, that focused on the potential impact of teaching on student learning based on principles of effective instruction. Jerome did not include an analysis of the student's action; instead he attended to the teacher and produced a claim about the opportunities afforded by the teacher's decisions.

On the other hand, it is important to recognize that analysis of students' responses and what can be inferred about students' thinking and understanding is a particularly complex skill. This analysis requires the combination of content knowledge and ability to analyze student

thinking (Ball, 1997; Hiebert et al, 2007; Santagata & Angelici, 2010). Without knowledge of the range of possible student responses, it is easy to overlook a response that counts as evidence. Second, even if PSTs attend to student responses, they must know what a particular response implies about student thinking (Bartell et al, 2012; Franke et al, 2007; Hiebert et al, 2007).

The majority of the participants had limited teaching experience. Furthermore, the teaching modeled at their fieldwork did not always align with ambitious practices. Therefore, PSTs had limited opportunities outside the methods course to observe student-centered teaching and the possible range of responses that arise when student thinking is elicited. Perhaps a claim based on principles of effective instruction constitutes an intermediary step towards a more sophisticated analysis.

These findings may thus contribute to our understanding of ways PSTs develop the ability to reason about teaching in integrated ways and of the challenges they may encounter. A plausible learning trajectory consists of three phases. PSTs first either attend to separate elements of teaching through descriptions, or, and this occurred often in our data and in other studies, they attempt to integrate student responses, math content, and teacher's goals and strategies but do so in problematic ways and arrive at unwarranted conclusions (Morris, 2006; Spitzer et al, 2010). As seen in the MMC group, PSTs may or may not develop the ability to use student-based evidence. Once a focus on students is introduced in teacher preparation, PSTs begin to consider student learning in their analyses (Santagata & Angelici, 2010; Davis, 2006; Stockero, 2008; van Es & Sherin, 2002). Initially principles of effective teaching for eliciting and building on student learning, as taught in their coursework, serve as lenses for viewing. When student-centered practices are present, PSTs notice and comment on them, explaining what the learning outcomes for students might be. Finally, PSTs begin to attend to and interpret

student learning based on evidence visible in the moment. To develop this last skill, structured opportunities to evaluate examples and non-examples of valid evidence seem necessary.

Another set of comments that merit discussion are those stating that more evidence was needed to be able to make claims about the effects of teaching on learning. Ten percent of the LMT responses fell in this category at posttest compared to 4% for the MMC group. This finding mirrors Stockero's results (2008): PSTs shifted from a definitive and often superficial analysis of student learning to a more tentative stance that considered multiple interpretations. Such stance is promising given that prior studies have shown this orientation and skill to be particularly challenging (Fernandez et al., 2003; Spitzer et al., 2010).

Our findings are promising but other potential contributing factors should be considered. One possibility is that the LMT group simply improved in their abilities to write about their reflections through various course activities. Findings would be more robust if additional measures were added, such as semi-structured interviews that investigate more deeply participants' reflective thoughts. The influence of field placements should also be considered. It is likely that field experiences interacted with methods course experiences and impacted PST's learning. Instead of distributing field setting variations randomly between groups, a follow-up study could examine the interplay between these two experiences and add valuable information for structuring fieldwork experiences and supporting mentor teachers.

While it is impossible to isolate one component of the course that was most helpful, literature on professional development suggest that perhaps core components are the *sustained* nature of the reflection that took place – the repeated opportunities to reflect on practice across different representations (e.g. video, student work, lesson plans) throughout two quarters — supported by a framework that facilitated *systematic* analysis of practice (Borko et al., 2008;

Givvin & Santagata, 2011; Seago, 2004; Stockero, 2008; van Es & Sherin, 2002). Consistent findings from other studies suggest that disciplined analysis of teaching must have a central role in teacher preparation if we want to foster future teachers' skills, knowledge and dispositions for sustained lifelong learning.

References

Ball, D. L. (1997). What do students know? Facing challenges of distance, context and desire in trying to hear children. In B. Bibble, T. Good, & I. Goodson (Eds.), *International handbook on teachers and teaching* (Vol. 2, pp. 679-718). Dordrecht, Netherlands: Kluwer Press.

Bartell, T. G., Webel, C., Bowen, B., Dyson, N. (2012). Prospective teacher learning: Recognizing evidence of conceptual understanding. *Journal of Mathematics Teacher Education.* Publish first on February 7, 2012. DOI: 10.1007/s10857-012-9205-4.

Berk, D., & Hiebert, J. (2009). Improving the mathematics preparation of elementary teachers, one lesson at a time. *Teachers and Teaching: Theory and Practice, 15*(3), 337-356.

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, *24*(2), 417-436.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markam (Eds.), *Handbook of child psychology*: (Vol. 3, pp. 77- 166). New York: Wiley.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively guided instruction*. Portsmouth, NH: Heinemann.

Carpenter, T.P., Franke, M.L., Levi, L. (2003). *Thinking mathematically: Integrating arithmetic & algebra in elementary school*. Portsmouth, NH: Heinemann.

Davis, E. (2006). Characterizing productive reflection among preservice elementary teachers: Seeing what matters. *Teaching and Teacher Education*, *22*, 281-301.

Empson, S. B., & Levi, L. (2011). *Extending children's mathematics: Fractions and decimals.*

Portsmouth, NH: Heinemann.

Fernandez, C., Cannon, J., & Chokshi, S. (2003). A US-Japan lesson study collaboration reveals critical lenses for examining practice. *Teaching and Teacher Education, 19*, 171–185.

Givvin, K., & Santagata, R. (2011). Toward a common language for discussing the features of effective professional development: The case of a U.S. mathematics program. *Professional Development in Education, 37,* 3, 439-451.

Hammerness, K., Darling-Hammond, L., Bransford, J., Berliner, D., Cochran-Smith, M., McDonald, M., & Zeichner, K. (2005). *How teachers learn and develop*. In L. Darling-Hammond, J. Bransford, P. LePage, K. Hammerness, & H. Duffy (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 358-389). San Francisco: Jossey-Bass.

Hawkins, D. (1973). What it means to teach. *Teachers College Record, 75*(1), 7-16.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K.C., Wearne, D., Murray, H., Plivier, A., Human, P. (2011). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Hiebert, J., Morris, A. K., & Glass, B. (2003). Learning to learn to teach: An "experiment" model for teaching and teacher preparation in mathematics. *Journal of Mathematics Teacher Education, 6*, 201-222.

Hiebert, J., Morris, A. K., Berk, D., & Jansen, A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education, 58*, 47-61.

Hiebert, J., & Stigler, J. W. (2000). A proposal for improving classroom teaching: Lessons from the TIMSS video study. *Elementary School Journal, 101,* 3-20.

Jacobs, V. R., & Philipp, R. A. (2004). Mathematical thinking: Helping prospective and

practicing teachers focus. *Teaching Children Mathematics, 11*, 194–201.

Jansen, A., & Spitzer, S. M. (2009). Prospective middle school mathematics teachers' reflective thinking skills: Descriptions of their students' thinking and interpretations of their teaching. *Journal of Mathematics Teacher Education, 12*(2), 133–151.

Kersting, N. (2008). Using video clips as item prompts to measure teachers' knowledge about teaching mathematics. *Educational and Psychological Measurement, 68*, 845-861.

Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning in and for practice. *Elementary School Journal, 109*, 491-509.

McDonald, M., Kazemi, E., Kavanagh, S.S. (2013). Core practices and pedagogies of teacher education: A call for a common language and collective activity. *Journal of Teacher Education, 64*(5), 378-386.

Moore-Russo, D. & Wilsey, J. N. (2014). Delving into the meaning of productive reflection: A study of future teachers' reflection on representations of teaching. *Teaching and Teacher Education, 37*, 76-90.

Morris, A. K. (2006). Assessing pre-service teachers' skills for analyzing teaching. *Journal of Teacher Education, 9*(5), 471-505.

Morris, A., Hiebert, J., & Spitzer, S. (2009). Mathematical knowledge for teaching in planning and evaluating instruction: What can preservice teachers learn? *Journal for Research in Mathematics Education, 40*(5), 491-529.

Santagata, R., & Angelici, G. (2010). Studying the impact of the lesson analysis framework on pre-service teachers' ability to reflect on videos of classroom teaching. *Journal of Teacher Education, 61*(4), 339–349.

Santagata, R., & van Es, E. (2010). Disciplined analysis of mathematics teaching as a routine of

practice. In J. W. Lott & J. Luebeck (Eds.), AMTE Monograph 7: Mathematics teaching: Putting research into practice at all levels (pp. 109–123). San Diego, CA: Association of Mathematics Teacher Educators.

Santagata, R., Zanoni, C., & Stigler, J. W. (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*(2), 123–140.

Seago, N. (2004). Using video as an object of inquiry for mathematics teaching and learning. In J. Brophy (Ed.), *Using video in teacher education* (pp. 249-286). San Diego, CA: Elsevier.

Stein, M.K. & Smith, M. (2011). *Five practices for orchestrating productive mathematics discussions*. Reston, VA: National Council of Teachers of Mathematics.

Spitzer, S. M., Phelps, C. M., Beyer, J. E. R., Johnson, D. Y., & Sieminski, E. M. (2010). Developing prospective elementary teachers' abilities to identify evidence of student mathematical achievement. *Journal of Mathematics Teacher Education. 14*(1), 67- 87.

Stockero, S. L. (2008). Using a video-based curriculum to develop a reflective stance in prospective mathematics teachers. *Journal of Mathematic Teaching Education, 11*(5), 373-394.

Van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2010). *Elementary and middle school mathematics: Teaching developmentally.* Boston, MA: Allyn & Bacon.

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*(4), 571-596.

Table 1

Summary of Key Analysis Skills, Instructional Resources, and Key Tasks for LMT Course

Key Analysis Skills	Instructional Resources	Key Tasks
Specifying learning goals	"Elementary and Middle School Mathematics" (Van de Walle et al., 2010) In-house student work and videos Online videos from Annenberg Learner website	Lesson planning Unit planning Unpacking of learning goals for lessons observed through video
Conducting evidence-based observations of student thinking and learning	"Cognitively Guided Instruction" (books and videos) (Carpenter et al., 1999; 2003) "Extending Children's Mathematics: Fractions and Decimals" (Empson & Levi, 2011) IMAP videos (Philipp et al., 2011) In-house student work and videos	Analysis of videotaped interviews of students solving math problems Interview with individual student and analysis of math understanding Analysis of videos of lesson taught by more experienced teachers, peers, and self Live observation and analysis of classroom lessons Analysis of student work collected from observed lessons
Reasoning about teaching and learning	"Making Sense: Teaching and Learning Mathematics" (Hiebert et al., 1997) "Five Practices for Orchestrating Productive Mathematics Discussions" (Smith & Stein, 2011) CGI videos (Carpenter et al., 1999; 2003) Online videos from Annenberg Learner website In-house videos	Discussion of videotaped lessons and teaching episodes Analysis of lessons taught by peers and self
Proposing improvements	In-house videos	Discussion of alternative teaching tasks or moves from what was observed in videotaped or live lessons

Table 2

Phase II Coding: Response Types When No Justified Claims Were Produced

Response Type	Definition	Example
Descriptive	Responses are purely descriptive, often reading like a summary of the events of the clip. There is no analysis of teacher action or student learning.	*When the student answers incorrectly, the teacher asks the student to explain his thinking. She then asks him another question. She never tells him if he is right or wrong.*
Missing relevant evidence	Responses produce a claim about the effects of teacher action on student learning but no evidence was cited or the evidence provided is not revealing of student learning.	*I like how the teacher said "tell me about your thinking", I also like how she helped him get the correct answer using multiple techniques such as sliding his finger, labeling the lengths, etc.*
Inaccurate depiction of mathematics	Responses inaccurately portray the mathematics at the core of the interaction between the teacher and the student(s).	*I think this is an effective way to teach how to compare fractions. It can be a difficult task for students especially with fractions that aren't very similar, but by showing this example, the student is able to visually see how they compare with each other.* [The response misinterprets the mathematical task in the clip. Students are not asked to compare fractions but to partition a whole into equal parts.]

Table 3

Phase II Coding: Response Types When Justified Claims Were Produced

Response Type	Definition	Example
Principles of effective instruction	Responses analyze a particular instructional decision or offer a suggestion that would provide an opportunity for student learning.	*The teacher asks the student to explain the best way to divide the number line to assess if the student is able to understand the basic concept of partitioning. She persists in asking the student's thinking to further assess his understanding and to make sure that he understands fractions on a number line so that he is not confusing tick marks versus the intervals in between the tick marks.*
Student learning	Responses include an analysis of the students' mathematics learning at a particular moment in the clip and the instructional move that may have supported students' learning.	*Students are using a length model to find the fractional parts of a whole. This student had the common misconception that in using a length model for fractions you would count the lines rather than the spaces. The teacher helped to clarify by asking the student to identify the parts of the race and [the student] soon realized that there were three parts instead of four.*
Request for more evidence	Response identifies a need for more evidence to make a valid claim about the effects of instruction on student thinking/learning.	*The teacher did a good job walking the student through the problem. The student seemed to understand "how" to do the procedure. It is still unclear if the student understands "why" he is doing what he is doing.*

Table 4

Types of Unjustified Claims Produced at Pre-Test for Each Group and Combined.

Response Type at Pretest	Number of Responses and Percentages		
	MMC	LMT	Combined
Descriptive	35 (41.2)	27 (33.3)	**62 (37.3)**
Missing relevant evidence	42 (49.4)	43 (53.1)	**85 (51.2)**
Inaccurate depiction of mathematics	8 (9.4)	11(13.6)	**19 (11.5)**
Total	**85 (100)**	**81 (100)**	**166 (100)**

Table 5

Types of Justified Claim Produced at Pre-Test for Each Group and Combined.

Response Type at Pretest	Number of Responses and Percentages		
	MMC	LMT	Combined
Student learning	14 (35.9)	11 (31.4)	**25 (33.8)**
Principles of effective instruction	21 (53.8)	18 (51.4)	**39 (52.7)**
Request for more evidence	4 (10.3)	6 (17.2)	**10 (13.5)**
Total	**39 (100)**	**35 (100)**	**74 (100)**

Table 6

Types of Unjustified Claim Responses for the MMC Group and LMT Group at Pre- and Post-Test.

Response Type	Number of Responses and Percentages			
	Pre MMC	Post MMC	Pre LMT	Post LMT
Descriptive	35 (41.2)	32 (43.2)	27 (33.3)	5 (19.2)
Missing relevant evidence	42 (49.4)	36 (48.7)	43 (53.1)	19 (73.1)
Inaccurate depiction of mathematics	8 (9.4)	6 (8.1)	11 (13.6)	2 (7.7)
Total	**85 (100)**	**74 (100)**	**81 (100)**	**26 (100)**

Table 7

Types of Justified Claim Responses for the MMC Group and the LMT Group at Pre- and Post-Test.

Response Type at Posttest	Number of Responses and Percentages			
	Pre MMC	Post MMC	Pre LMT	Post LMT
Student learning	14 (35.9)	19 (38.0)	11 (31.4)	32 (35.6)
Principles of effective instruction	21 (53.8)	26 (52.0)	18 (51.4)	46 (51.1)
Request for more evidence	4 (10.3)	5 (10.0)	6 (17.2)	12 (13.3)
Total	**39 (100)**	**50 (100)**	**35 (100)**	**90 (100)**

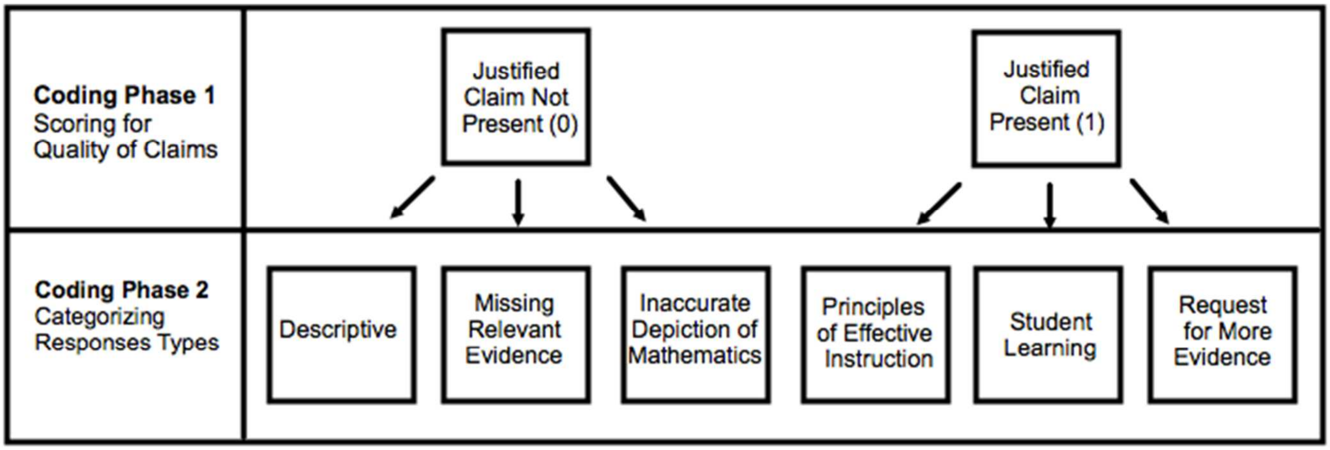

Figure 1. Coding Tree

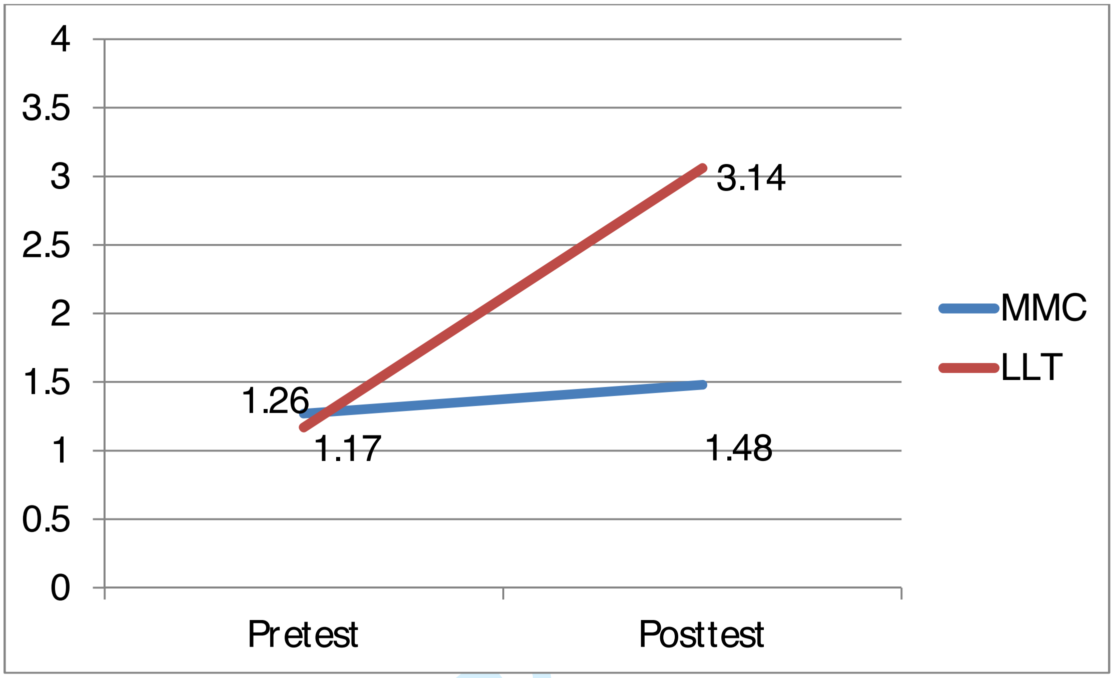

Figure 2. Typical Methods Course (MMC) and Learning from Teaching (LMT) Group Performance on Pre- and Post-test for Quality of Hypothesis Produced.