Integrating Mathematics and Pedagogy: An Investigation of the Effects on Elementary Preservice Teachers' Beliefs and Learning of Mathematics[1]

Theoretical Framework

Our work grows out of two arenas: mathematics teacher education, and beliefs.

Mathematics Teacher Education

Most elementary school children in the United States are not developing mathematical proficiency. One concern has been that teachers' mathematical content knowledge may not be deep enough to support the kind of rich mathematical teaching called for in Standards documents. Historically, the development of prospective teachers' content knowledge takes place in undergraduate courses years before prospective teachers study methodology. We agree with the rationale for this sequence: One cannot teach what one does not know, and so a prospective teacher must understand mathematics to be in a position to teach it. However, we find problematic the idea's translation into practice, so that prospective elementary school teachers first learn mathematics content and only later consider issues of teaching and learning. First, we know that developing deep understanding of the mathematics of elementary school is far more difficult than was once thought (Ball, 1990; Ma, 1999; Sowder, Philipp, Armstrong, & Schappelle, 1998). Second, our experience has been that even when prospective elementary school teachers (PSTs) attend a thoughtfully planned course designed to engage them in rich mathematical thinking, too many of them go through the course in a perfunctory manner. Many PSTs' expectations about what mathematics is—a fixed set of rules and procedures—along with their perceptions of how children and adults learn mathematics—by being shown how to solve problems in a prescribed step-by-step fashion—clash with the more conceptual, meaning-making goals mathematics-course designers hold for PSTs. Thus, in what we believe is a better place to begin work with prospective elementary school teachers, we have implemented and studied a new model for integrating mathematical content and children's mathematical thinking: Prospective elementary school teachers (PSTs), while

[1] Preparation of this paper was supported by a grant from the National Science Foundation (NSF) (REC-9979902). The views expressed are those of the authors and do not necessarily reflect the views of NSF.

enrolled in their first mathematics course for elementary school teachers, Math 210, engage with children's mathematical thinking years before they begin student teaching.

Many PSTs report having had bad experiences learning mathematics (Ball, 1990). They do, however, care about children (Darling-Hammond & Sclan, 1996). We believe that we can facilitate the process by which PSTs come to learn mathematics by beginning with that about which PSTs care most—children (see Figure 1). We place children (rather than children's thinking, for example) at the center of caring because we believe that for most PSTs, the initial caring is a phenomenological act of concern for the whole child versus for a particular characteristic of the child. For example, following is a PST's written response regarding what she found most valuable in a course in which she learned about children's mathematical thinking while working with children:

> I think it was just amazing. I had never ever taken a class like this one, so I am very surprised and happy that I took it. The most valuable (I think) is the contact with the kids. (Marisol, 12/5/01)

Tapping into PSTs' caring about children is the first step, but we hypothesize that when PSTs engage children in mathematical problem solving, the PSTs' circles of caring expand to include children's mathematical thinking. For example, the PST who responded below indicated that she not only found her work with children valuable but also appreciated the chance to understand their mathematical thinking:

> This experience with working with the children, itself, is what I think makes this class valuable. Also, I think that analyzing the way children solve problems is valuable information for us to reference to when we, finally, become teachers, especially to become more effective in teaching mathematics. (Isabel, 12/4/01)

When working with children, PSTs begin to see how children think about mathematics and come to recognize that children solve problems in varied and sometimes mathematically powerful ways. We predict that at that time many PSTs' circles of caring will extend to mathematics, because they realize that to be prepared to understand the depth and variety in children's mathematical thinking, they must themselves grapple with the mathematics. Figure 1 shows a model that captures our view of the growth of PSTs' caring.

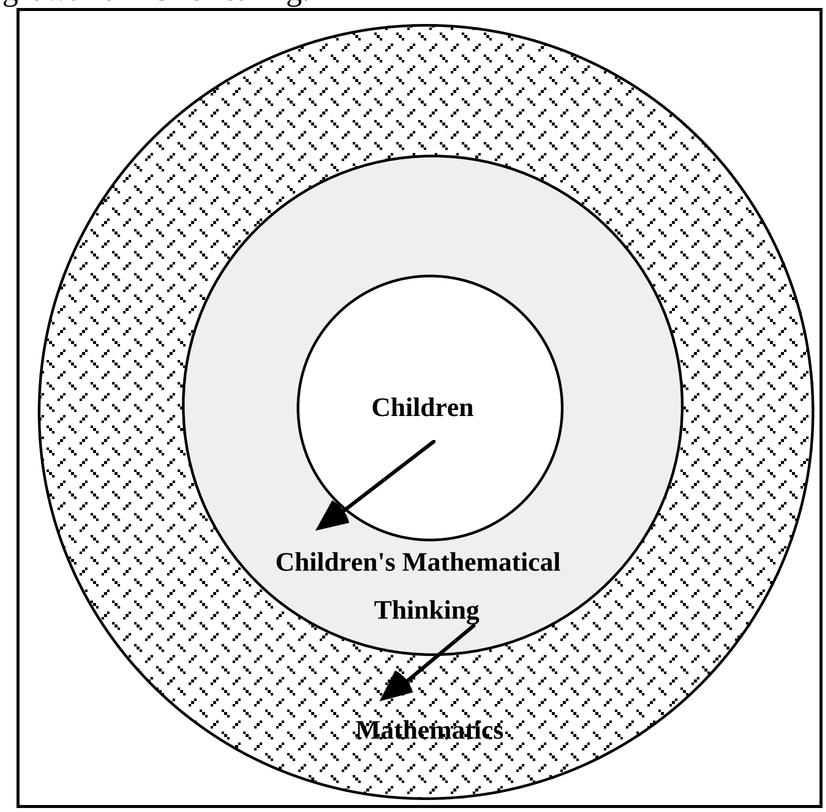

Figure 1. A model of growth, by way of children's mathematical thinking, from PSTs' caring about children to their caring about mathematics.

This is an old idea. More than 100 years ago John Dewey (1990) wrote that every subject has two aspects, "one for the scientist as a scientist; the other for the teacher as a teacher" (p. 351). He wrote, "[The teacher] is concerned, not with the subject-matter as such, but with the subject-matter as a related factor in a total and growing experience [of the child]. Thus to see it is to psychologize it" (p. 352). In our approach to working with PSTs, we assume that they can come to care (Noddings, 1984), not as scientists, but as teachers, about mathematics.

Beliefs

Beliefs and content knowledge are intertwined. Although to separate these two is difficult, we have tried to do so. The term *belief* is so common in education literature today that many who write about beliefs do so without defining the term and instead assume that researchers know what beliefs are (Thompson, 1992). We identify four components of beliefs that are important for the way we attempt to measure beliefs. First, beliefs influence perception (Pajares, 1992). That is, beliefs serve to filter some complexity of a situation to make it comprehensible, and, therefore, when inferring beliefs, one must determine to what one attends in a situation. Second, beliefs are not all-or-nothing entities; they are, instead, held with different intensities (Pajares, citing Rokeach, 1968); thus, when measuring beliefs, we consider tasks that offer multiple interpretation points. Third, beliefs tend to be context specific (Cooney, Shealey, & Arvold, 1998), and, hence, we situate beliefs items in particular contexts and infer a respondent's beliefs on the basis of his or her interpretation of the context. Fourth, beliefs might be thought of as dispositions toward action (Cooney et al.,1998; Rokeach, 1968); therefore, we infer one's beliefs on the basis of how the person might act in a particular situation.

However one chooses to define (or not define) *beliefs*, "for the purposes of investigation, they must be inferred" (Pajares, 1992, p.315). For our work, we needed a beliefs instrument to administer to large numbers of prospective elementary school teachers years before they were in the classroom. We identified three problems with using Likert Scales and attempted to overcome them with our instrument. (Figure 2 lists two items drawn from Likert-scale beliefs instruments.) First, knowing how a respondent interprets the words used in items is difficult. For example, for Item 2 (see Figure 2), one needs to know whether the respondent distinguishes among situations in which the child listens: when teachers demonstrate procedures, when teachers present problem situations, and when students share unusual thinking. For Likert items, respondents are asked to agree or disagree with statements, whereas in our survey, respondents use their own words to react to, or answer questions about, learning situations. Although this format does not remove the need to draw inferences, it reduces it.

Second, we think that beliefs can be inferred by determining to what one attends in a complex situation, and Likert scales seldom provide contexts. For example, in Item 1 (see Figure 2), would whether one imagined a kindergarten class or an 8th-grade algebra class be relevant? In our instrument, each item is embedded in a context, so one can better determine to what the respondent's attention was drawn.

Item 1. In mathematics, perhaps more than in other fields, one can find set routines and procedures. (Collier, 1972)

Item 2. It is important for a child to be a good listener in order to learn how to do mathematics. (Fennema, Carpenter, Loef, 1990)

Figure 2. Two Likert items.

Third, Likert items do not carry with them good ways for assessing the depth with which one holds a belief. One may respond in a way that indicates the existence of a belief that is not central to the respondent. McGuire (1969) stated, "When asked, people are usually willing to give an opinion even on matters about which they have never previously thought" (p. 151). For example, a respondent may agree strongly with Item 2 (see Figure 2) but not believe that listening matters as much as speaking or other activities in mathematics. We addressed this issue in our beliefs survey by drawing inferences from that to which respondents attended in learning episodes and when they attended to certain issues.

After determining that we would not use Likert Scales to assess beliefs, we next had to finalize our choice of beliefs to assess. To that end, we asked ourselves and several colleagues, "If you could identify one belief that, if held by prospective elementary school teachers (PSTs) entering your mathematics course, might increase the likelihood that the PSTs would be poised to benefit from what you had to offer, what belief would you change?" Also, drawing upon the work of Franke, Carpenter, Levi, and Fennema (2001), we were interested in beliefs that might be called *generative* "because prospective teachers who develop them will continue to grow in their learning of mathematics and will continue to develop beliefs that will help them to implement the reforms articulated in the *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics, 2000) once they begin to teach" (Ambrose, 2003, p. 10). We identified 7 beliefs listed in Appendix 1 (we initially identified 8 beliefs, but our piloting indicated that we could not reliably assess one of the beliefs [In case it comes up in a question, that belief was: “It takes a variety of episodes for most children to learn. That is, just because a student has been exposed does not mean he or she has learned.”]).

Research Design

We randomly assigned 159 PSTs enrolled in Math 210, the first of four mathematics courses for prospective elementary school teachers, to one of five treatments. The treatments are

Children’s Mathematical Thinking Experience-Live (CMTE-L) (*n* = 50)

Children’s Mathematical Thinking Experience-Video (CMTE-V) (*n* = 27)

Mathematical Observation and Reflection Experience-Reform (MORE-R) (*n* = 23)

Mathematical Observation and Reflection Experience-Traditional (MORE-T) (*n* = 25)

Control (*n* = 34)

The CMTEs

The CMTE-L and CMTE-V were 2-unit courses that met for 14 sessions over the semester. Each CMTE-V class session met for 2 1/2 hours whereas the CMTE-L class sessions were a little shorter, only 2 1/4 hours, to enable students to drive back to campus in time for their next classes. Both courses might be thought of as concurrent labs for Math 210, although the instructors of the CMTE-L and CMTE-V courses did not collaborate with the Math 210 instructors. Note, the Math 210 content was whole number and rational number concepts and procedures, and the Math 210 instructors were graduate students working toward master’s degrees in mathematics. These instructors received instructional support from a senior mathematics educator. The instructors for the CMTE courses were mathematics educators. The focus of the CMTE courses was mathematics from the child’s perspective, and extensive use of video clips created to highlight students’ mathematical strengths and weaknesses were used in both courses. For example, we use one video clip to highlight a difficulty children experience related to fractions. This video clip is used in our courses, and it is also illustrative because it was taken during an interview between a CMTE student and a child, and the follow-up discussion provides a taste for what took place in our CMTE courses. [Show video of Jackie_Fractions] Specific issues we hoped our video clips might address included children's invented strategies, children's agility with numbers, difficult-to-follow children's reasoning, and children's mistakes and misconceptions. Overall we hoped to provide opportunities for PSTs to see their own need for developing deeper understanding of the mathematics of elementary school so that they could make sense of children's rich mathematical reasoning and support children who clearly had not had opportunities to make sense of mathematics.

The CMTE-L students worked with children on eight occasions, interviewing and tutoring them. The CMTE-V students did not work with children, but in addition to viewing video clips developed by our project that were also used by students in the CMTE-L, the CMTE-V students viewed "real-life" video clips of interviews conducted by students in the CMTE-L class. That is, on about seven occasions, the CMTE-V students received unedited video clips of entire interviews conducted the previous day by CMTE-L students working with children. To provide some privacy to those PSTs interviewing children on video, we showed only the child, without showing the PST interviewer in these videos.

The MOREs

PSTs assigned to the MOREs made 14 weekly visits to local elementary schools selected on the basis of their proximity to campus and the availability and willingness of teachers to participate. Pairs of PSTs were assigned to teachers, according to the PSTs' schedules and transportation capabilities and the teachers' mathematics class time. PSTs made one 90-minute visit, at least 45 minutes of which was intended to be during mathematics instruction, per week to a local elementary classroom. At midsemester PSTs were assigned new classrooms and grade levels, so that over the course of the semester they made 7 visits at primary grades and 7 visits at intermediate grades. Each week the PSTs in the MOREs wrote a 1–2 page reflection paper about the visit. Below are questions designed to help them focus their reflection papers:

- What does the teacher do?

- What are the students doing? What might they be thinking or reasoning about? What opportunities for interaction or active engagement occur?
- Did anything surprise/intrigue you today?
- Were important mathematics introduced, developed, or practiced during the lesson? If so, what?
- Were questions used to promote understanding and facilitate student thinking?
- What questions do you have about the lesson?

They PSTs also wrote mid- and end-of-semester reflections about the experience, addressing the following questions:

- What have you learned from the last 7 visits?
- Have these visits affected your thinking about becoming an elementary school teacher? If so, in what ways?
- Have these visits affected your thinking about mathematics instruction? If so, in what ways?
- Would you recommend these experiences to a friend who wants to become a teacher? Why or why not?
- Have these visits affected your experience in or your thinking about Math 210?

Our general goals for assigning PSTs to the MORE-T were to provide PSTs with opportunities to make connections between their university courses and the world of teaching by acquainting them with what takes place in schools and classrooms, by providing them with opportunities to work with students of diverse backgrounds, and by providing a sense of mathematics curricula and mathematics instruction typically used in schools today. The teachers were provided latitude about how the PSTs should spend their time in class, and some of the teachers in the MORE-T classes arranged for PSTs to help children with their mathematics work. Because we expected that preservice teachers who observed qualitatively different approaches to mathematics instruction would come to views of mathematics different from those of preservice teachers who observed more traditional mathematics lessons, we assigned one group of PSTs (MORE-T) to observe teachers we considered *traditional* and another group (MORE-R) to observe teachers we considered to be *reform* oriented. Our general goals for assigning PSTs to the MORE-R were to provide PSTs opportunities to view the variety of mathematical strategies that children can construct when encouraged to do so, to convince preservice teachers that reform-teaching practices are effective, and to acquaint preservice teachers with examples of reform mathematics curricula currently being used in some schools.

Teachers were selected on the basis of lesson observations, interviews, and recommendations. Typically, the instruction in the reform classrooms was qualitatively different from that in most traditional classrooms. The reform teachers taught conceptually based lessons using reform materials or programs and encouraged student discourse and use of multiple strategies while they attempted to promote student understanding by employing problem solving and explorations. Reform teachers also tended to spend more time on mathematics daily.

Control Group

The students assigned to the control group were enrolled in Math 210 and completed the instruments administered to all PSTs in the study but were offered no additional experience. To entice enough students to participate in our study, we paid students assigned to the CMTEs or MOREs up to $600 for their semester's work, including completing the surveys. Control students were paid only to complete the surveys, but because they were paid $24/hour, many earned more than $100 by completing the pre and post surveys.

Instruments

To investigate the treatments' effects on our students' knowledge of mathematics and beliefs about mathematics, teaching, and learning, all participants completed a beliefs assessment and a content assessment as pretests and posttests. In addition, we conducted select individual and group interviews at various points during the study. The instruments were developed by our research team and are described below.

IMAP Web-Based Beliefs Survey

Survey development. We developed an instrument to assess beliefs that might affect PSTs' subsequent learning of mathematics: beliefs about mathematics and mathematics understanding and learning. To conduct a large-scale, randomized study with quantitative data, we needed an instrument from which we could derive a common metric for measuring change in individuals and for comparing individuals with one another. We also wanted the instrument to provide qualitative data that could be used for more holisitic analysis. To avoid the Likert-scale limitations, outlined above, we developed a web-based instrument, using video and teaching episodes about which PSTs construct responses (instead of choose from options provided).

The beliefs instrument and the accompanying scoring rubrics were developed over a 2-year period by project researchers with support from other staff members. We used a recursive cycle of development that included piloting segments of the instrument, analyzing PSTs' responses to the segments, revising the segments, and piloting them again.

The instrument contains seven segments. Each segment includes several questions about a particular situation. Four segments are in the domain of whole number; two are in the domain of fractions; and one is a more general teaching segment. The chosen domains were the domains of focus for our experimental treatments and were important topics in the mathematics-for-teachers course in which the PSTs were enrolled. Two segments included video clips of individual children solving mathematics problems with an interviewer. Each segment is associated with two or three beliefs, and each belief is assessed using a separate rubric for each of two or three segments. Overall, we developed 17 rubrics for the instrument.

To illustrate how we assigned scores for each belief, we describe one segment, the rubric used to assign scores to prospective teachers' responses to that segment, and the scoring system used to combine scores on individual rubrics to determine an overall score for the belief. The scores on segments and on beliefs reflect the amount of evidence of holding that belief a respondent provided. This scoring is in keeping with the idea that beliefs can be held with different intensities and are more or less central (Rokeach, 1968). In one segment, prospective teachers watch a video clip of a teacher presenting a story problem to a 6-year-old child in a one-on-one setting: "There are 20 kids going on a field trip. Four children fit in each car. How many cars do we need to take all 20 kids on the field trip?" After a long pause, the child stated 10 as his answer. He confirmed that he had guessed when the teacher asked. The teacher then directed the child to show her the kids by counting out 20 cubes. She then reminded him that 4 kids fit in one car and asked him to show her 4 kids in one car. She directed him to make another group of 4 for the next car, and he followed her directions. She continued in this fashion until he had made five groups of 4 cubes. She then reminded him that each group stood for a car and prompted him to count each of the cars. She counted along with him.

Our interpretation of this clip was that the teacher was overly directive and focused the child's attention on counting cubes instead of on understanding the relationships among the quantities in the situation. She could have provided prompts that were less specific, to see whether the child could solve the problem with less help. For example, she might have invited him to try to use the cubes to represent the situation, waiting to see what he would do before providing additional help. The clip featured a familiar real-world context and manipulatives. The teacher was encouraging in her tone of voice and in providing the child with praise. These positive features of the clip were quite attractive to some respondents, leading them to focus on these aspects instead of on the excessive guidance offered by the teacher, as is evident in the following response:

> I thought it was good that she let him try and answer the problem first and then she showed him how to figure it out using the blocks. ... They need to test things out themselves and then see the different ways to approach a problem. … I think the strengths of this video were allowing the child to think on his own and solve the problem. … I didn't see any weaknesses in this video clip. I really liked it.

In addition to being impressed with the teacher's use of blocks, this respondent wrote about the importance of letting the child figure out the problem for himself. She used the rhetoric that we would like PSTs to employ, but in this case she applied the rhetoric in a context in which we believe it was inappropriate. Responses like this one reminded us of the critical role that context played in inferring beliefs from responses. Without knowing the context to which these comments were directed, one might interpret this response as providing strong evidence of Belief 7: During interactions related to the learning of mathematics, the teacher should allow the children to do as much of the thinking as possible.

We were particularly concerned about whether the respondents noted that the teacher did too much leading and, if so, when they noted that fact. Those who noted excessive guidance in their response to the first prompt, "Please write your reaction to the video clip. Did anything stand out for you?" provided strong evidence of this belief because the issue mattered so much to these respondents that it shaped their interpretation of the episode. For subsequent prompts, "Identify the strengths of the teaching in the episode" and "Identify the weaknesses of the teaching in the episode," some respondents noted, after the third prompt, that the teacher might have provided too much guidance. In such cases, we determined that the respondents provided some evidence of the belief. It was not strong evidence because the issue did not matter enough to them to shape their initial interpretation of the clip. Because the survey was web-based, respondents could not change earlier responses. Open-ended questions allowed us to discern which issues mattered enough to respondents to affect their interpretations. (Note. Survey Development copied and pasted mostly verbatim from Ambrose et al., 2003).

Rubric development. For rubric development, a major task, two research teams met two to four times per week in 3–4-hour sessions for 6 months. To get a wide range of responses for each item, we began with data gathered from three groups: prospective-teacher participants in a pilot of an experimental treatment (pre and post data), expert mathematics educators, and mathematics education graduate students. We later gathered responses from prospective teachers in a second pilot of the treatment; thus, our rubric development was based on a set of about 80 responses.

We adopted a grounded-theory approach (Glaser & Strauss, 1967), using pilot data, to develop each rubric. Initially, each rubric-development-team member independently analyzed the entire set of responses to a particular item with a particular belief in mind and sorted the responses into categories. Those responses that provided the greatest evidence of the belief were placed into one category, whereas those that provided no evidence of the belief were placed into another category. The remaining responses were placed into one, two, or three groups, depending on how each team member categorized responses. To determine the appropriate category for each response, the team members looked for degrees of evidence of the belief in question. Team members met to discuss and reach consensus on the number of categories and those responses that belonged together in each category. The scoring teams then created descriptions of each category. We needed descriptions that were both robust, describing aspects of the belief that the responses provided, and procedural and concrete, so that others using the rubrics could code the responses with a high degree of reliability. This rubric-negotiation process was lengthy, and the development process took approximately 72 person-hours per rubric (4 weeks x 6 hours per week x average of three people per team). Sometimes negotiations concerned the number of categories, whereas in other cases, negotiations concerned the descriptions of the categories. We often traversed the terrain from the theoretical to the practical. We described categories to one another, then reanalyzed the data to check that the descriptions provided the glue that held the category together with regard to the belief in question. Once we had reached consensus on a rubric's categories and descriptions, we reanalyzed the data to check for inter- and intra-rater reliability. We then shared the rubric with the other team to test for coherence, reliability, and validity, a critical component of our work. The other team's members used the rubrics to code the data; we then compared the development group's codes with the testing group's codes to determine interrater reliability.

To determine a final score for each belief, we combined individual scores from each rubric. We developed a rubric-of-rubrics system that could be applied to each belief. (Note. Rubric Development section was copied and pasted mostly verbatim from Clement et al., 2003).

Coding the study data. Mathematics education graduate students, external to the project, (most from other universities), were trained to code the large-scale beliefs data. IMAP project researchers trained the coders but had no say in final codes assigned by the coders. The responses were blinded so that neither coders nor trainers could determine in which treatment the respondents were enrolled or whether the responses came from presurveys or postsurveys. Twenty percent of the responses were double-coded, and we achieved, on average, 84% reliability on the 17 rubrics.

IMAP Mathematics Content Assessment

We designed a content assessment to determine whether the treatments had measurable effects on the college students' performances on items that address the main content of the accompanying mathematics course (MATH 210): place value and rational numbers (both fractions and decimals). Items for this special instrument could well have been used for the final examination, and indeed a few of the items were selected from previous MATH 210 final examinations. The focus of the items was on conceptual understanding rather than on computational skill (Figure 3 shows an item from the content assessment).

5. Antonio asks, "When I multiply [for example, 49×23, shown to the right], why do I have to put in the zero [points to the zero in 980]?"

What would you say to Antonio?

$$\begin{array}{r} 49 \\ \times\ \underline{23} \\ 147 \\ \underline{980} \\ 1127 \end{array}$$

Figure 3. One content-assessment item.

We piloted preliminary versions of the content assessment. These results were used in two ways. One use was to gauge the clarity and likely value of the items, as judged by the students' written responses and by an occasional debriefing of a student who had completed the test. The second use was to suggest rubrics with which to score the items calling for explanations. These rubrics were composed by the content-test team in preparation for the large-scale study described in this paper.

Administrations and scoring. Used as a pretest and a posttest, the instrument was scored by four pairs of paid scorers qualified to exercise the judgments needed for scoring with rubrics. The responses were blinded so scorers knew neither from what treatment a response came nor whether it was a pretest or posttest. Each day, one of the content-test team trained each pair of scorers on the rubrics for one page of the content assessment, explaining the rubrics before the scorers mutually scored five tests. The scorers next further refined their training through scoring another 20 tests to reach a high degree of cross-scorer reliability. Each scorer then marked a set of 30 tests, with a random 10 in common with the set of the other scorer. The two scorings of the common 10 were then compared, with a target of at least 80% agreement as an index of reliability. Most reliability-agreement percentages were in the 90%–100% range. A sample rubric, for the item in Figure 3, is shown in Figure 4.

Weighting the items. The rubric scores are probably, at best, ordinal scores, with a higher score representing a judgment of a higher degree of conceptual focus in the respondent's answer. Rubrics for some items are on a 0–1–2 scale, with those for others on 0–1–2–3 or 0–1–2–3–4 scales. Furthermore, the items are not equally *important*, just as items on an examination might not be weighted equally. Hence, we (a) rescaled the rubrics to have a more uniform scale across rubrics and also (b) weighted the items so that item scores could be added together in a meaningful fashion, much as one does in weighting items on examinations. Accordingly a panel of four mathematics educators met several times to rescale the rubrics for uniformity and to assign weights to each item. Objectively scored items were usually weighted 1; rubric-scored items were usually weighted 2; some objectively scored items were weighted 2 because correct responses were deemed to require considerable reasoning; and three items (one dealing with importance of the unit in fraction work and two requiring the writing of a story problem for a given computation) were adjudged by the panel to be particularly important for teachers and were weighted 3. After these rescalings and weightings, the maximum possible score for the content assessment (as augmented by the content items from the beliefs assessment) was 82.

Question 5. Antonio asks, "Why is the 0 in the multiplication algorithm?" (scored 0–2)

Explanation score	Examples
Explanation score 2 algorithmic with very good understanding OR simply good understanding; ▪ uses place-value language with good explanation; ▪ says something to the effect that we are multiplying 20 times 49 and not 2 times 49; states that the 98 is really 980 with good reason.	• When you got the 980 you were multiplying 20 times 49—however it's easier to think of 2 x 49 and just add zero to make it 20. • I would say you have to put the zero in because otherwise the numbers wouldn't be in the correct place. When you are multiplying 49 times 2, the 2 is actually 20 so the zero needs to go in for correct place value. • The 2 is in the tens place, the 3 is in the ones place. You first multiply 49 x 3; that gets you 147. Now you do the same for the 2, but since the 2 is in the tens place, it's the same as 49 x 20; that's why you bring down the 0.
Explanation score 1 uses place value language or names the places mechanically, NOT clearly stated that 2 times is really 20 times, OR states 20 but then "messes up." In this category the 0 is seen as a place holder for the ones place; a reference made to consider the 3 in 23; states that the 98 is really 980 without "good" reason	• Because you need something to replace the ones place. • Because that zero (ref. to a zero that he drew over the 3 in 23) is taking place of the three. • You put the zero there to show that you are not adding any numbers in the ones place; the zero allows you to know that even though there are no ones, that place value is still there, acting like there is so you don't put numbers in their wrong place values. • You have to take into consideration the 3 in 23. • Ok, the reason you put a 0 is because you are moving from ones place (1) to the tens place (10); the difference between 1 and 10 is the zero. • Now we are working the tenths [*sic*] place rather than the ones place; therefore you need to move the value over 1 and add a 0. • Because the #s we are multiplying are in the 10ths [*sic*] place (the #2), and we have to account for that by adding the zero so we know the 98 is from the 10ths [*sic*] place.
Explanation score 0: no understanding, nonsense, solely algorithmic; says only place holder	• Because every time you finish one row of numbers and move to the next, you have to put a 0 first because then the next number can be under that row of numbers. • Just refers to "place holder," as in "0 is a placeholder."

Figure 4. The rubric for the content item on placing the zero in multidigit multiplication.

Analysis and Results

Belief Data

Quantitative Belief Data

Each of the seven beliefs was treated independently, and for each belief all 10 pairwise comparisons among the four treatments and control were conducted, with the Holm's procedure used to reduce the likelihood of a Type I error. (In using the Holm's procedure, one sequences the p-values in increasing order and compares the smallest p value to $.05/n$; if the result is significant, one then compares the second smallest p-value to $.05/(n – 1)$, and so on. Once a result is not significant, then all results with larger p values are also considered nonsignificant. In using the Holm's procedure in our study with 10 comparisons for each belief, we compared the smallest p value to .005 (.05 / 10), the second smallest p value to 0.005556 (.05/9), and so on.) The belief data were treated as ordinal, not interval, data; thus, distribution-dependent statistical tests, such as ANOVAs, were inappropriate. Pairwise differences between groups were analyzed as follows:

1. Individual subjects were assigned change scores representing either no positive change, a small positive change, or a large positive change between their presurvey and postsurvey scores. Each subject received one change score for each of the seven beliefs.
2. Change scores were analyzed using a polychotomous log-linear odds ratio using the ordered logit procedure in the STATA software package. An ordered logit model depends on the natural log of the odds ratio between two variables. A value greater than 1 for an odds ratio between two groups indicates that a member of one group is more likely than a member of the other to be in the next-higher category of the dependent variable (in our case, one of the change-score categories). An ordered logit procedure generates a log-linear regression and corresponding goodness-of-fit statistic to test the prediction that change scores vary by group assignment. (Because some readers may not have experience with odds ratios, we provide a simple example using only two categories per group: Suppose that 9 of 10 men applying to a university are accepted whereas 8 of 10 women are accepted. How might one compare the

number of men accepted to the number of women accepted? One could state that there are 9/8 times as many male applicants admitted as female applicants. One could also state that there are twice as many female applicants denied admission as male applicants. These two statements, though both true, leave one with a different sense. How is one to deal with this difference? One solution is to standardize the comparison by use of an odds ratio. To do so, one composes a ratio of those *in* to those *not in* for each group and then compares these ratios. In this example, the ratio of males' acceptances to males' rejections is 9:1, whereas the ratio of females' acceptances to females' rejections is 8:2, or 4:1. Or, another way to put this is that for every male rejected, 9 are admitted, whereas for every female rejected, 4 are admitted. The odds ratio is a means for capturing all this information; it is created by taking the ratio of the odds, which in this case would be 9:1 / 8:2, or 9/4, or 2 1/4. One interprets this odds ratio by noting that the odds of acceptance among males is 2 1/4 times the odds of acceptance among females. This does *not* mean 2 1/4 times as many males are admitted as females, nor does it mean that males are 2 1/4 times as likely to be admitted as females. It means that for every male rejected, the number of males accepted is 2 1/4 times the number of females accepted for every female rejected.)

Belief-data results. Although 31 of the 70 pairwise comparisons of beliefs resulted in a p value of less than .05, only 18 of these 31 tests were significant when the p-values were shared, using the Holm's procedure. We now discuss these 18 statistically significant differences. In the next section we examine the descriptive data to determine what additional trends might be inferred from the belief data.

Table 1 shows the significant differences on pairwise belief comparisons, and Tables 2–8 (located at the end of the paper) list, for the seven beliefs assessed, the percentages of students, by treatment, in each of the three change-score categories.

Table 1
Significant Differences on Change Scores, by Belief

Group	**Belief**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
CMTE L	More-T Control	More-T Control	Control	More-T	More-R More-T Control	More-T	
CMTE-V		More-T Control	More-T Control	More-T Control		More-T	More-T

For each significant difference, the belief scores of students in the CMTE-L or of students in the CMTE-V increased more than the belief scores of students in one of the other three groups. CMTE-L students' belief scores improved significantly more than scores of the control students on four of the seven beliefs, significantly more than MORE-T students' scores on five of the seven beliefs, and significantly more than MORE-R students' scores on one of the seven beliefs. CMTE-V students' belief scores improved significantly more than control students' scores on three of the seven beliefs and significantly more than MORE-T students' scores on five of the seven beliefs. No statistically significant differences were found among any other pairwise comparisons: Students in the MORE-R, MORE-T, and control groups did not score significantly higher than one another on any beliefs. Furthermore, no significant differences were found between CMTE-L and CMTE-V students' scores. Evidently, participating in either of the CMTEs led to greater belief change for PSTs than participating in the MORE-T or the control group.

Nonsignificant but noteworthy differences. Large-scale, randomized assignment of subjects to treatments and quantitative and qualitative analyses of data, as carried out in this study, are expensive to employ in educational environments, so we discuss some nonsignificant differences to highlight observations that might be worth considering but that would be expensive to replicate. Although only one significant difference was found related to the MORE-R (CMTE-L vs. MORE-R on Belief 6), the CMTE-L and CMTE-V groups' score increases were greater than those of the control group, the MORE-R, and the MORE-T on every belief. We examined differences between change scores for the CMTE-L and CMTE-V groups and found no patterns. In other words, although learning about children's mathematical thinking leads to more belief change than not learning about children's mathematical thinking, we could detect no differences in belief change between PSTs who learned about children's mathematical thinking by a combination of viewing video clips and working with children and those who learned about children's mathematical thinking only by viewing video clips.

An important issue with implications for teacher education is the effect of assigning PSTs to visit traditional classes. The MORE-T group's scores improved about the same amount as the control group's on Beliefs 2, 3, and 5, and the MORE-T group's were slightly favored over the control group's on Belief 1. However, for Beliefs 4, 6, and 7, the control group's scores improved more than those of the PSTs in the MORE-T, and the change differences for each of these three beliefs were larger than for the other beliefs. Without finding significant differences, one can draw no clear conclusions, but the results of this study raise concerns about the possible negative effects on PSTs' beliefs of assigning them to visit traditional classrooms.

Qualitative Belief Data

To provide the reader (and presession participants) with a sense of the qualitatively different ways individuals in different treatments responded to items, we report specific belief-survey responses for one item used to measure Belief 6 *(The ways children think about mathematics are generally different from the ways most adults would expect them to think about mathematics. For example, real-world contexts, manipulatives, and drawings support children's initial thinking whereas symbols often do not)* and two PSTs' pretest and posttest responses to the item (see Figure 5 for the item).

8.1 Place the following four problems in rank order of difficulty for children, and explain your ordering (you may rank two or more items as being of equal difficulty). Easiest =1.

a) Understand 1/5 + 1/8	Select rank	Please explain your rank
b) Understand 1/5 X 1/8	Select rank	Please explain your rank
c) Which fraction is larger1/5 or 1/8, or are they same size?	Select rank	Please explain your rank
d) Your friend Jake attends a birthday party at which five guests equally share a very large chocolate bar for dessert. You attend a different birthday party at which eight guests equally share a chocolate bar exactly the same size as the chocolate bar shared at the party Jake attended. Did Jake get more candy bar, did you get more candy bar, or did you and Jake each get the same amount of candy bar?	Select rank	Please explain your rank

8.2 Which of these two items did you rank as easier for children?
___c is easier than d ___d is easier than c ____both items are equally difficult
Please explain your answer.

Figure 5. Segments 8.1 and 8.2.

This item was designed to assess whether respondents recognized that real-world contexts often support children's mathematical thinking, whereas symbols (in this case, fraction symbols) are often confusing for children. Differences between the CMTE-Ls' and MORE-Ts' scores and those between the CMTE-Vs' and the MORE-Ts' scores were statistically significant on this belief. The pretest and posttest responses from one student in the CMTE-V and one student in the MORE-T are provided in Tables 9 and 10. The CMTE-V student ranks the word problem as most difficult of the four items (rank of 4) in the pretest but as easiest (rank of 1) in the posttest. In the posttest, this student wrote about the difficulties students might have understanding what the symbols mean when no real-world context is provided. Compare this response with that of a student in the MORE-T group who ranked the word problem as difficult (rank of 3, tied with the compare problem) in the pretest, then as most difficult (rank of 4) in the posttest. Additionally, in his posttest comments, the MORE-T student specifically mentioned the difficulty the students with whom he worked (in his MORE-T placements) had with word problems. In other words, whereas the CMTE-V student moved from viewing symbols as easier for children than real-life contexts to seeing real-life contexts as easier for children than symbols, the MORE-T student's belief as measured by this item moved in the opposite direction. The rubric score for this item was one of three rubric scores aggregated in the score for Belief 6. Table 7 **(**at the end of the paper**)** and Figure 6 show, for each treatment, the percentages of students whose Belief 6 change (pretest to posttest) scores were categorized as large increases, small increases, and no increases.

Table 9
Presurvey and Postsurvey Responses From a PST in the Children's Mathematical Thinking Experience—Video (CMTE-V) Treatment

Treatment	Rationale for rank of Item c, Which fraction is larger, 1/5 or 1/8?	Rationale for rank of Item d, Comparing 1/5 and 1/8, in context	Easier	Explanation of choice of easier item
CMTE-V Presurvey	Rank = 1 (easiest) Easier because by knowing that 8 is bigger than 5 they will see that 1/5 is larger than 1/8	Rank = 4 (hardest) word problems are harder for children to understand and, this problem involves several things not just a simple adding or multiplying problems. Several steps need to be taken to solve it.	C	As I explained before, comparing fractions is easier than solving a word problem.
CMTE-V Postsurvey	Rank = 4 (hardest) I think children have a harder time comparing fractions. The child may automatically assume that 1/8 is bigger because 8 is bigger than 5.	Rank = 1 (easiest) its a real world problem and its something they can probably relate to...sharing candy bars. Children may not know much about fractions, but they know about who gets bigger shares.	D	as explained before. the problem involves sharing and it is put in a real world context, it is easier for children to see the problem in this way.

Table 10
Presurvey and Postsurvey Responses From a PST in the Mathematical Observation and Reflection Experience in a Traditional Classroom (MORE-T) Treatment

Treatment	Rationale for rank of Item c, Which fraction is larger, 1/5 or 1/8?	Rationale for rank of Item d, Comparing 1/5 and 1/8, in context	Easier	Explanation of choice of easier item
MORE-T Presurvey	Rank = 3 Once multiplying is learned them you can teach him/her about common denominators and cross multiplying and such.	Rank = 3 Use same methods to see what fraction is bigger or smaller.	C	C is easier than D because the problem is already set out in front of you, where as D you have to go find the numbers and make sure you put them in the right place.
MORE-T Postsurvey	Rank = 2 have to know that a fifth is bigger that an eighth	Rank = 4 (hardest) Word problems always difficult for children, lots of places to get tripped up on.	C	Just due to the fact that I see more children stumble on word problems because they don t know what info is important and what is not.

How might we explain these data? The students in the CMTE-V watched video clips of children making sense of problems situated in relevant contexts. For example, they watched a child incorrectly solve 4 – 1/8 represented symbolically, but the same child correctly solved the problem situated in the context of having 4 cookies and eating one eighth of one cookie. On the other hand, the students visiting the classrooms may not have had the same opportunities to observe children working with relevant real-life story problems. Perhaps the children at the placements of the MORE-T students were solving more traditional but unrealistic word problems often found in textbooks, problems that may have held little meaning for the children asked to solve them. One might infer from this result that placement in a traditional classroom early in PSTs' teacher training might actually interfere with the PSTs developing the beliefs that we would like for them to develop. The data in Table 7 and Figure 6 support this conjecture. Although no significant pairwise differences were found in the Belief 6 score changes of the students in the MORE-T and those in the control group, the percentage of students who showed no increase pretest to posttest on Belief 6 was higher in the MORE-T group (68%) than in the control group

(50%), whereas the percentage of students who showed a large increase from pretest to posttest on Belief 6 in the control group (16%) was higher than in the MORE-T (0%).

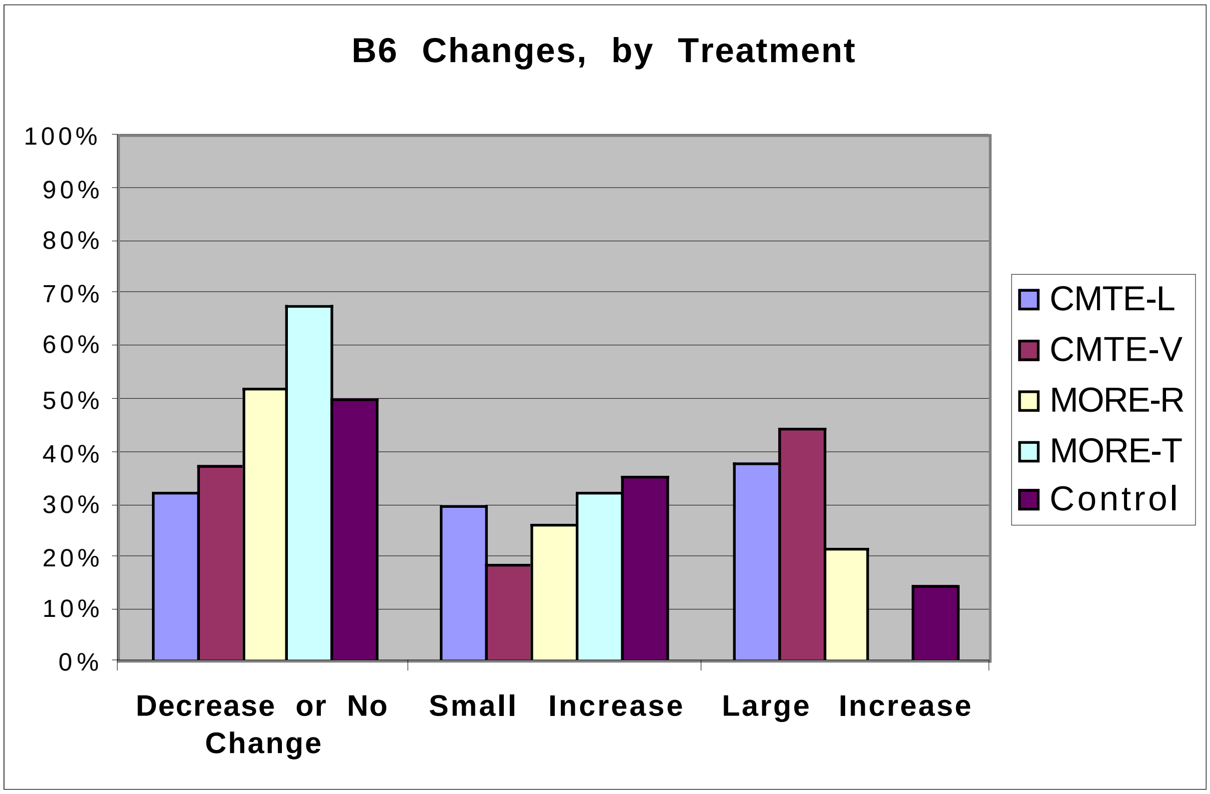

Figure 6. Belief 6 score changes, by treatment.

Content Data

The content data were treated as interval data, and all 10 pairwise comparisons among the four treatments and the control group were conducted using *t*-tests. Holm's procedure was used to reduce the likelihood of a Type I error.

Quantitative Content Data

Because for some pairwise comparisons we could not justify a prediction in a certain direction, we analyzed all pairwise comparisons using two-tailed tests. Table 11 shows the pretest, posttest, and change scores by treatment on the content test. Although the average change scores were higher for the CMTE groups than for the other groups, no pairwise comparisons were significant.

Table 11
Pretest, Posttest, and Change-Score Averages (With Standard Deviations) by Treatment

Treatment	Pretest average	Posttest average	Average change score (Standard deviation)
Control (*n* = 34)	37.6	50.4	12.8 (8.08)
CMTE-L (*n* = 50)	36.8	51.5	14.7 (7.64)
CMTE-V (*n* = 27)	39.1	52.6	13.5 (9.29)
MORE-R (*n* = 23)	38.7	49.9	11.2 (7.98)
MORE-T (*n* = 25)	32.3	44.6	12.3 (7.69)

Note. Content test has 82 possible points.

We also compared all those students who studied children's mathematical thinking (CMTE-L and CMTE-V students) with those who did not study children's mathematical thinking (MORE-T, MORE-R, and control students) using a *t*-test. Because we expected that those who studied the mathematics embedded in children's mathematical thinking developed richer mathematical understanding than those who did not, this comparison was treated as a one-tailed test. Table 12 shows the pretest, posttest, and change scores by these pooled groups. Results of a *t*-test show that these differences were significant at the .05 level. The effect size was 0.26.

Table 12
Pretest, Posttest, and Change-Score Averages (With Standard Deviations) for CMTE Students and for Control and MOREs Students

Treatment	Pretest average	Posttest average	Average change score (Standard deviation)
CMTEs (*n* = 77)	37.6	51.9	14.3 (7.87)
Control/MOREs (*n* = 82)	36.3	48.5	12.2 (8.21)

Note. Content test has 82 possible points.

Every treatment has students who are not poised to benefit from their classes, for reasons that had little to do with the treatments. For example, more than 5% of the students, including at least one student from each treatment, scored lower on the posttest than on the pretest for reasons that could not be explained by a pretest ceiling effect. Table 13 shows the percentages of students by treatment whose scores increased at least the stipulated number of points between the pretest and the posttest. Looking at the outliers may not be interesting, because every class has special students on each end of the spectrum. For example, the CMTE-V group had both the highest percentage of students whose scores increased by at least 25 points and of those whose scores decreased. Therefore, we focus on the bulk of the students in the middle, that is, those whose scores increased more than 10 points, more than 15 points, and more than 20 points.

TABLE 13
Percentages, by Treatment, of Students Whose Scores Increased at Least the Stipulated Numbers of Points From Content Pretest to Posttest

	< 0	≥ 0	≥ 5	≥ 10	≥ 15	≥ 20	≥ 25	≥ 30
MORE-R	4.3%	95.7%	87.0%	47.8%	21.7%	8.7%	4.3%	4.3%
MORE-T	4.0%	96.0%	88.0%	52.0%	28.0%	24.0%	8.0%	0.0%
CMTE-V	11.1%	88.9%	81.5%	59.3%	51.9%	29.6%	11.1%	0.0%
CMTE-L	2.0%	98.0%	88.0%	68.0%	50.0%	28.0%	6.0%	0.0%
Control	5.9%	94.1%	88.2%	52.9%	38.2%	23.5%	2.9%	2.9%
Total	5.03%	94.97%	86.79%	57.86%	40.25%	23.89%	6.27%	1.24%

In each of these three middle categories, the highest percentages of students were the CMTE-L and the CMTE-V students. For example, 2/3 of the CMTE-L and 3/5 of the CMTE-V students' scores, more than in the other treatments, increased at least 10 points. Consider those who increased at least 15 points: Half the CMTE students' scores (51.9% for CMTE-V students and 50% for CMTE-L students) increased at least 15 points, compared with scores of 38.2% of the control students, 28% of the MORE-T students, and 21.7% of the MORE-R students. These percentages drop for increases greater than 20 points, but again the CMTEs had the highest percentage of students with such score increases. We investigated whether the cut-offs chosen were responsible for these results by calculating the percentages of students who increased at least 12.5 points (26.1% MORE-R, 40.0% MORE-T, 51.9% CMTE-V, 60.0% CMTE-L, and 47.1% control) and 17.5 points (17.4% MORE-R, 24.0% MORE-T, 40.7% CMTE-V, 44.0% CMTE-L, and 26.5% control), but these data are also compelling.

You may wonder why control students seem to outperform students in the MORE-R and MORE-T. We suspect that because most students in the four treatments completed all posttests but almost half of the control students did not return for the posttests, a high proportion of those control students who did not complete the posttests may have been weaker students. (The percentages of those completing both content and belief pretests who returned for the posttests were 94% for CMTE-L, 90% for CMTE-V, 89% for MORE-T, 77% for MORE-R, and 56% for control.) The monetary incentive to take the posttests was greater for the treatment students than for the control students because we withheld the second half of their semester stipend, up to $300,until they completed all the instruments at the end of the semester. We withheld nothing from the control students because they had already been paid for completing the pretests; thus, the only financial incentive the control students had to return for the posttests was the pay they would receive for the time spent completing the posttests. If this hypothesis is correct, the result is an underreporting of differences due to the CMTEs. That is, had we found a way to keep all the control students in the study (for example, by paying them more money to complete the posttests), the differences might have been even greater. (By the way, we recommend a relatively simple solution to this problem. Instead of offering the students $24/hour to complete the two posttests, we could have offered $50/hour up to $200 for completing the two posttests. We believe this rate might have resulted in more control students' completing the posttest.)

Qualitative Content Data

How the CMTEs supported PSTs' mathematics learning. At the end of the semester we met with a group of students drawn from all the treatments to discuss the Math 210 course and its relationship to their other experiences. At that time, several students in the CMTEs spoke about how their work with children supported their learning in the Mathematics 210 course. Meg spoke about how the CMTE-V helped her develop deeper understanding of the "simple concepts," when she stated, "I just know that before I took this class I was in Math 210 and I was just freaking out. I was thinking that I'm going to have to teach this to kids and I don't understand this. ... But then from going to the night [CMTE-V] class and getting a deeper understanding on so many simple concepts, it helped a lot."

Kerry spoke about how the CMTE-V course helped reinforce what she was learning in Math 210 and how, when she solved mathematics problems, she thought back to her observations of children making sense of mathematics:

> I think it reinforced it, for me. It was like, I learned it. I learned it on the blackboard and on the paper in [Math] 210. And then I went to 296 [CMTE-V], and we did it. And I saw kids doing it. And I was like, 'Oh, yeah.' And so you really ... you remember. ... I'm starting to remember more now. Like when I look at a math problem—like when I was doing the surveys again at the end, I was like thinking of it how I relearned it in 296 [CMTE-V]. Like how I learned it on paper in [Math] 210, and then how I learned it like with my hands, and with seeing other kids do it in 296 [CMTE-V].

Heidi found that the CMTE-V caused her to think more deeply about the mathematics she was learning: "First of all, I did learn about math. Because sadly, I knew how to do it. But the concepts, the ... seriously, I was learning as much as the kids were learning. And it was *so* beneficial to me. But above all, just how they thought about math." Heidi remembered a video clip of a fifth-grade girl who reflected on the fact that although her teacher usually taught by expecting her students to make sense of the mathematics, the teacher taught one lesson by explaining a procedure, and the student did not remember it because she had not figured it out for herself. Heidi said, "Like the first thing that will probably always stick in my mind is that little girl [who said], 'I didn't figure it out myself.' And like [speaking for myself], 'Yeah. I didn't figure it out for myself, either. That's why I don't get it."

This notion of thinking deeply about more than the procedures came through from Phil, a student in the CMTE-L course.

Phil: One thing I got out of 296 [CMTE-L]—if I hadn't taken 296, I probably would have gone through 211, 312, 313 [the subsequent mathematics courses] focusing on the thing that I already knew, the algorithm that I already knew, and thinking, "All right, that's the best." But now I realize that I have to take it all in, everything that the class is teaching, not just what I think is the most important. Because all of this is important. I probably wouldn't have realized that if I hadn't taken 296.

RP: Why is it important?

Phil: Because people think in different ways, and not everyone thinks like me.

RP: And you don't think you would have gotten that from [Math 210 alone?

Phil: No. No way.

How Does the CMTE Help?

How do students benefit from working with children? We were unable, through our beliefs assessment, to measure any differences in belief change between those who studied children's mathematical thinking using only video and those who used video in combination with working with children. However, anecdotally, students who worked with children very much appreciated the experiences. CMTE-L students' responses to an open-ended survey at the end of the class indicated that for 94% of them, the CMTE-L had affected their experiences in the content course; 55% of all PSTs attributed the effect to *their interactions with children*, and 40% suggested that the CMTE-L provided a rationale for learning the content in the mathematics course. Pat's response is an example:

> Taking the [CMTE-L] has enabled me to make a practical application to what I learned inside my Math 210, and use it and see what function it serves in the elementary classroom. (12/4/01)

In answer to the question "Specifically what did you think was valuable or not valuable in the Math 296 course?" half the students mentioned that they valued working with the children; 37% stated that they valued the opportunity to learn about children's mathematical thinking; and 39% stated that they valued learning how to teach mathematics. For these students, the CMTE-L course seems to have been something between an experience for learning mathematics and an experience for learning about teaching and learning. Heather, a student in the CMTE-L course, captured the more pedagogical benefits of the CMTE-L in her comments about the class:

> I remember my parents saying, "You're taking a math class [meaning an extra mathematics class, the CMTE-L]. No. Take art. Take something fun." I was like no. I told them, "It is fine." It was the most fun class. My semester would have been horrible without—I know that sounds totally horrible—but it seriously was a way for me to do something that was towards my major. It definitely, out of all the classes I took this semester, I think I learned the most about myself. I

> learned about the way I think and I know that what I learned in this class will stick with me through my teaching and through my other teaching classes. It has just given me a different outlook.

Optimal placement of a CMTE experience. To investigate the relationship between when a CMTE-L course is offered in a PSTs' undergraduate experience and what PSTs gain from the experience, we taught a CMTE-L to students nearing their entry into a credential program. We selected the students from those enrolled in Math 313, the last of the four mathematics courses for Liberal Studies majors at our university. Math 313 is described as a capstone course for prospective elementary school teachers; it addresses algebra, number systems, transformational geometry, and problem solving. Generally, a focus in this course is on quantitative reasoning embedded in story problems, especially those involving multiplicative reasoning; in this work, students review issues of fractions, decimals, and percents. The last third of class is qualitative algebra, with attention to slope.

We recruited students enrolled in Math 313 to serve in our treatment and control groups. The treatment group was small because schedules of students nearing graduation are constrained, for both academic and personal reasons. Ten of the 12 students enrolled in the CMTE-L completed the pretests and the posttests; 22 of the 47 control students who completed the pretests also completed the posttests. In our analysis, we compared the 10 CMTE-L students with the 22 control students.

Our Project Team spent much time considering and consulting others on how we should treat our belief data, and, as mentioned, in our primary analysis, we chose a conservative stance and treated the data as ordinal; hence, we applied the odds ratio. We also investigated treating the data as interval data and used *t*-tests; then we compared the *t*-test results to the odd-ratios results (see Table 14). The two analyses gave the same results in 66 of the 70 pairwise comparisons, with 17 tests resulting in significant differences using either test and 49 results showing no significant differences using either test. In just four cells the results differed: in three we found significant differences for the *t*-test but not for the odds ratio, and for one we found a significant difference for the odds ratio but not for the *t*-test. On the basis of these results, we have decided to treat the data as interval for purposes of comparing the 313 data with the 210 data.

Table 14
Significant Pairwise Differences, by Belief, Using Odds Ratio (O) and T-Tests (T)

Treatment groups	B1	B2	B3	B4	B5	B6	B7
Control vs. CMTE-L	O, T	O, T	O, T	T[1]	O, T		
Control vs. CMTE-V		O, T	O, T	O, T			
Control vs. MORE-T							
Control vs. MORE-R							
CMTE-L vs. CMTE-V							
CMTE-L vs. MORE-T	O, T	O, T		O, T	O, T	O, T	T[1]
CMTE-L vs. MORE-R					O, T		
CMTE-V vs. MORE-T		O, T	O, T	O, T		O[1]	O, T
CMTE-V vs. MORE-R				T[1]			
MORE-T vs. MORE-R							

[1]Cells in which results of the two tests differ.

Table 15 shows these data. We calculated the average pretest and posttest belief scores for the 313 control group (Rows 1 and 3), the 313 CMTE-L group (Rows 2 and 4), the 210 control group (Rows 8 and 10), and the 210 CMTE-L group (Rows 9 and 11). We then calculated the pretest to posttest differences for the 313 control group (Row 5), the 313 CMTE-L group (Row 6), the 210 control group (Row 12), and the 210 CMTE-L group (Row 13). We then calculated, for each belief, the difference of these differences, so that Row 7 shows how much more than the 313 control students' belief scores the 313 CMTE-L students' belief scores increased; Row 14 shows how much more than the 210 control students' belief scores the 210 CMTE-L students' belief scores increased. We made several observations. First, the total average belief changes, if taking such averages means anything, are identical. That is, both CMTE-L groups increased their belief scores, on average, 2/3 of a point more than the corresponding control groups. However, note that the increases were not in the same beliefs. Whereas the 210 CMTE-L students' scores showed greater increases over their control counterparts' scores for Beliefs 1, 2, and 3, the 313 CMTE-L students' scores showed greater increases over their control counterparts' scores for Beliefs 5, 6, and 7. These differences are captured in Row 15, which shows the differences of the differences (Row 7 minus Row 14).

Table 15
Differences Between (313 CMTE-L and Control Pre-Post Differences) and (210 CMTE-L and Control Pre-Post Differences) by Belief

Row		B1	B2	B3	B4	B5	B6	B7	Total
1	313 Pretest Control (n=22)	0.96	1.00	1.96	0.96	1.35	1.57	0.61	**1.20**
2	313 Pretest CMTE-L(n=10)	1.00	1.20	1.70	1.20	1.40	1.80	0.50	**1.26**
3	313 Posttest Control (n=22)	1.04	1.35	2.00	1.04	1.48	1.83	0.61	**1.34**
4	313 Posttest CMTE-L(n=10)	1.50	2.20	2.30	1.80	2.60	2.90	1.20	**2.07**
5	313Pre to Post Control Difference	0.09	0.35	0.04	0.09	0.13	0.26	0.00	**0.14**
6	313 Pre to Post CMTE-L Difference	0.50	1.00	0.60	0.60	1.20	1.10	0.70	**0.81**
7	313 Diff of Diff	0.41	0.65	0.56	0.51	1.07	.84	.70	**0.67**
8	210 Pretest Control (n = 34)	0.47	0.35	0.68	0.18	0.88	0.74	0.29	**0.51**
9	210 Pretest CMTE-L (n =50)	0.36	0.34	0.64	0.32	0.74	0.90	0.36	**0.52**
10	210 Posttest Control (n = 34)	0.97	0.65	1.12	0.68	1.41	1.18	0.85	**0.98**
11	210 Posttest CMTE-L (n=50)	1.60	1.70	1.94	1.38	1.96	1.96	1.14	**1.67**
12	210Pre to Post Control Difference	0.50	0.29	0.44	0.50	0.53	0.44	0.56	**0.47**
13	210Pre to Post CMTE-L Difference	1.24	1.36	1.70	1.94	1.38	1.96	1.96	**1.14**
14	210 Diff of Diff	0.74	1.07	0.86	0.56	0.69	0.62	0.22	**0.68**
15	(313 diff of diff) – (210 diff of diff)	**-.33**	**-.42**	**-.30**	**-.05**	**0.38**	**0.22**	**0.48**	**0.00**

What do these data indicate? Inasmuch as Beliefs 1, 2, and 3 are about mathematics and Beliefs 5, 6, and 7 are about learning and teaching mathematics, we note that working with children seemed to help students early in their undergraduate studies change their beliefs about mathematics more than their beliefs about teaching and learning; however, working with children later in their undergraduate experiences affected students' beliefs about learning and teaching more than their beliefs about mathematics. Thus, the decision about where in one's teacher-preparation program the CMTE experience should be included may depend upon the reasons for including it at all.

Implications

If providing opportunities for prospective elementary school teachers to learn about children's mathematical thinking early in their undergraduate experiences is helpful, the next issues to consider are how to offer more such experiences and what factors play out in faculty's ability to offer such a course. Currently many mathematics methodology instructors offer such experiences, but how might these experiences be infused earlier? At SDSU we have developed a children's mathematical thinking experience that is currently being required of all Liberal Studies majors, and local community colleges are also beginning to offer this course. We think, however, that a promising way to infuse children's mathematical thinking into prospective elementary school teachers' undergraduate experiences early enough to affect their concurrent and subsequent learning of mathematics is to infuse children's mathematical thinking into mathematics content courses. Reactions to experiments with this infusion, with faculty both at SDSU and around the country, have been positive. This endeavor has raised questions about how faculty use video clips of children's thinking and to attain what goals, but those are questions for a subsequent study. We end with two quotes from faculty who express enthusiasm for including such video clips in their mathematics content courses:

> I have used the tape to show my prospective elementary teachers the kind of creative and "different" thinking students use to reason and make calculations. The video clips became motivational clips and saved me having to make the argument for PUFM (Profound Understanding of Fundamental Mathematics, Ma,1999). (George Poole, personal communication, November 12, 2001)

> I used the video you provided last summer with my content course—the first about 6–7 parts. It had a wonderful effect. One student just remained sitting after class when others were leaving. She finally said, "This is my last math class [we require two]. I don't think I know enough math to teach these students we saw!" (Mary Ann Lee, personal communication, March 12, 2002)

References

Ambrose, R. C., Philipp, R. A, Chauvot, J., & Clement, L. L. (2003). *A web-based survey to assess prospective elementary school teachers' beliefs about mathematics and mathematics learning: An alternative to Likert scales*. Manuscript submitted for publication.

Ball, D. L. (1990). Breaking with experience in learning to teach mathematics: The role of a preservice methods course. *For the Learning of Mathematics, 10*(2), 10–16.

Clement, L. L., Chauvot, J., Philipp, R. A, & Ambrose, R. C. (2003). *A method for developing rubrics for research purposes.* Manuscript submitted for publication.

Collier, C. P. (1972). Prospective elementary teachers' intensity and ambivalence of beliefs about mathematics and mathematics instruction. *Journal for Research in Mathematics Education, 3*, 155–163.

Cooney, T. J., Shealy, B. E., & Arvold, B. (1998). Conceptualizing belief structures of preservice secondary mathematics teachers. *Journal for Research in Mathematics Education, 29*, 306–333.

Darling-Hammond, L., & Sclan, E. M. (1996). Who teaches and why: Dilemmas of building a profession for twenty-first century schools. In J. Sikula (Ed.), *Handbook of research on teacher education* (2nd ed., pp. 67–101). New York: Simon & Schuster Macmillan.

Dewey, J. (1990). *The school and society: The child and the curriculum*. Chicago: The University of Chicago Press.

Fennema, E., Carpenter, T. P., & Loef, M. (1990). *Mathematics beliefs scales.* University of Wisconsin-Madison.

Franke, M. L., Carpenter, T. P., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal, 38*, 653–690.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine: New York.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahweh, NJ: Erlbaum.

McGuire, W. J. (1969). The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (pp. 136–314). Reading, MA: Addison-Wesley.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–332.

Rokeach, M. (1968). *Beliefs, attitudes, and values: A theory of organization and change.* San Francisco: Jossey-Bass.

Sowder, J. T., Philipp, R. A., Armstrong, B. E., & Schappelle, B. P. (1998). *Middle-grade teachers' mathematical knowledge and Its relationship to instruction*. New York: SUNY Press.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127–146). New York: Macmillan.

Beliefs

Beliefs about mathematics

1) Mathematics, including school mathematics, is a web of interrelated concepts and procedures.

Beliefs about knowing or learning mathematics or both

2) One can perform standard algorithms without understanding the underlying concepts

3) Understanding mathematical concepts is more powerful and more generative than remembering mathematical procedures.

4) If students learn mathematical concepts before they learn standard algorithms, they are more likely to understand the algorithms when they learn them. If they learn the algorithms first, they are less likely ever to learn the concepts.

Beliefs about children (students) doing and learning mathematics

5) Children can solve problems in novel ways before being taught how to solve such problems. Children in primary grades generally understand more mathematics and have more flexible solution strategies than their teachers, or even their parents, expect.

6) The ways children think about mathematics are generally different from the ways adults would expect them to think about mathematics. For example, real-world contexts, manipulatives, and drawings support children's initial thinking whereas symbols often do not.

7) During interactions related to the learning of mathematics, the teacher should allow the children to do as much of the thinking as possible.

Tables

Table 2

Belief 1 Score Change (Pretest to Posttest) by Treatment : Percentages (Numbers) of Students in Each Change Category

Group	No change or decrease	Small increase	Large increase
CMTE-L (*n* = 50)	18% (9)	46% (23)	36% (18)
CMTE-V (*n* = 27)	30% (8)	41% (11)	30% (8)
MORE-R (*n* = 23)	39% (9)	39% (9)	22% (5)
MORE-T (*n* = 25)	44% (11)	52% (13)	4% (1)
Control (*n* = 34)	59% (20)	26% (9)	15% (5)

Table 3

Belief 2 Score Change (Pretest to Posttest) by Treatment : Percentages (Numbers) of Students in Each Change Category

Group	No change or decrease	Small increase	Large increase
CMTE-L (*n* = 50)	30% (15)	22% (11)	48% (24)
CMTE-V (*n* = 27)	33% (9)	26% (7)	41% (11)
MORE-R (*n* = 23)	48% (11)	26% (6)	26% (6)
MORE-T (*n* = 25)	72% (18)	12% (3)	16% (4)
Control (*n* = 34)	71% (24)	18% (6)	12% (4)

Table 4

Belief 3 Score Change (Pretest to Posttest) by Treatment : Percentages (Numbers) of Students in Each Change Category

Group	No change or decrease	Small increase	Large increase
CMTE-L (*n* = 50)	32% (16)	22% (11)	46% (23)
CMTE-V (*n* = 27)	22% (6)	22% (6)	56% (15)
MORE-R (*n* = 23)	35% (8)	30% (7)	35% (8)
MORE-T (*n* = 25)	60% (15)	16% (4)	24% (6)
Control (*n* = 34)	65% (22)	21% (7)	15% (5)

Table 5

Belief 4 Score Change (Pretest to Posttest) by Treatment : Percentages (Numbers) of Students in Each Change Category

Group	No change or decrease	Small increase	Large increase
CMTE-L (*n* = 50)	32% (16)	36% (18)	32% (16)
CMTE-V (*n* = 27)	22% (6)	40% (11)	37% (10)
MORE-R (*n* = 23)	57% (13)	26% (6)	17% (4)
MORE-T (*n* = 25)	80% (20)	20% (5)	0% (0)
Control (*n* = 34)	53% (18)	35% (12)	12% (4)

Table 6

Belief 5 Score Change (Pretest to Posttest) by Treatment : Percentages (Numbers) of Students in Each Change Category

Group	No change or decrease	Small increase	Large increase
CMTE-L (*n* = 50)	22% (11)	38% (19)	40% (20)
CMTE-V (*n* = 27)	41% (11)	33% (9)	26% (7)
MORE-R (*n* = 23)	70% (16)	26% (6)	4% (1)
MORE-T (*n* = 25)	60% (15)	36% (9)	4% (1)
Control (*n* = 34)	50% (17)	38% (13)	12% (4)

Table 7

Belief 6 Score Change (Pretest to Posttest) by Treatment : Percentages (Numbers) of Students in Each Change Category

Group	No change or decrease	Small increase	Large increase
CMTE-L (*n* = 50)	32% (16)	30% (15)	38% (19)
CMTE-V (*n* = 27)	37% (10)	19% (5)	44% (12)
MORE-R (*n* = 23)	52% (12)	26% (6)	22% (5)
MORE-T (*n* = 25)	68% (17)	32% (8)	0% (0)
Control (*n* = 34)	50% (17)	35% (12)	15% (5)

Table 8

Belief 7 Score Change (Pretest to Posttest) by Treatment : Percentages (Numbers) of Students in Each Change Category

Group	No change or decrease	Small increase	Large increase
CMTE-L (*n* = 50)	54% (27)	16% (8)	30% (15)
CMTE-V (*n* = 27)	33% (9)	48% (13)	19% (5)
MORE-R (*n* = 23)	74% (17)	22% (5)	4% (1)
MORE-T (*n* = 25)	80% (20)	20% (5)	0% (0)
Control (*n* = 34)	47% (16)	41% (14)	12% (4)