

A WEB-BASED SURVEY TO ASSESS PROSPECTIVE ELEMENTARY SCHOOL TEACHERS' BELIEFS ABOUT MATHEMATICS AND MATHEMATICS LEARNING: AN ALTERNATIVE TO LIKERT SCALES¹

Rebecca Ambrose, University of California at Davis

Randolph Philipp, Jennifer Chauvot, and Lisa Clement

San Diego State University

Because of the importance of beliefs, mathematics educators need to consider ways to assess beliefs and belief change. Beliefs, because they must be inferred, can be difficult to measure, particularly with a common metric that enables one to compare individuals. Some limitations of Likert scales are identified and overcome in a newly developed instrument in which prospective teachers are provided scenarios to interpret. The instrument captures qualitative data that are quantified for purposes of comparison. Results from two administrations of the instrument demonstrate that it is an effective tool for assessing belief change.

Imagine that prospective elementary school teachers (PSTs) are entering a mathematics course and that you could change one of their beliefs to increase the likelihood that the PSTs would be poised to benefit from what the instructor had to offer? What belief would you change? None of several mathematicians and mathematics educators to whom we have posed this question has questioned its implicit underlying assumption, that beliefs can make such a difference. We take these responses as confirming evidence of the important role beliefs play in mathematics teaching and learning. In considering this question ourselves, we identified seven beliefs we would like to cultivate in PSTs; three are listed in Figure 1.

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

Belief 3 (About knowing or learning mathematics or both). Understanding mathematical concepts is more powerful and more generative than remembering mathematical procedures.

Belief 7 (About children [students] doing and learning mathematics). During interactions related to the learning of mathematics, the teacher should allow the children to do as much of the thinking as possible.

Figure 1. Beliefs of interest.

Given the importance of beliefs, assessing beliefs and belief change would benefit mathematics educators. As part of our large-scale research project, Integrating Mathematics and Pedagogy (IMAP), we needed an instrument to assess the belief change

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.

that we expected as the result of treatments presented concurrently with prospective teachers' first mathematics course for teachers. We first explain why we considered Likert-scale instruments insufficient for assessing the belief change of interest. We then describe the web-based belief instrument we developed and implemented and provide a rationale for its benefits over Likert-scale instruments. Finally, we provide data that indicate that our survey is general enough to capture a range of positions on beliefs but sensitive enough to capture change.

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 belief items in particular contexts and infer a respondent's belief 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 belief on the basis of how the person might act in a particular situation.

Why Likert Surveys Are Insufficient for Our Purposes

However one chooses to define (or not define) beliefs, "for the purposes of investigation, they must be inferred" (Pajares, 1992, p.315). Our work required us to identify a belief instrument that might be administered to large numbers of prospective elementary school teachers years before they were in the classroom.

We identified three problems with Likert Scales and attempted to overcome them with our instrument. (Figure 2 lists two items drawn from Likert-scale belief 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 determining how important the issue is to the respondent. One may respond in a way that may indicate 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 belief survey by drawing inferences from that to which respondents attended in learning episodes and when they attended to certain issues.

Our Belief Instrument

We set out to create an instrument to assess beliefs that might affect PSTs’ subsequent learning of mathematics: beliefs about mathematics and mathematics understanding and learning. We wanted an instrument that would provide a common metric for measuring change in individuals and for comparing individuals to one another. We also wanted the instrument to provide qualitative data that could be used for more holistic analysis. To avoid the limitations of Likert scales, outlined above, we developed an instrument in which PSTs construct responses instead of choosing from options provided. We later quantified these constructed responses using rubrics. We designed items to measure only the seven beliefs we had identified (three of the seven are listed in Figure 1).

Development of Instrument

This instrument and the accompanying scoring rubrics were developed over a 2-year period by the authors 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 doing 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 of the segments, 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 a respondent provided related to the belief. 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 the 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 and then waited to see what he would do before providing him with 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.

The complete rubric used to score this segment (for Belief 7) is provided in Figure 3. We were particularly concerned about whether the respondents noted that the teacher did too much leading and when they noted that fact. Those that 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 that the teacher might have provided too much guidance after the third prompt. In this case, 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.

No Evidence Overall satisfaction with guidance provided by teacher No teaching weaknesses identified	No Evidence Thought the teacher should explain more.
Weak Evidence In initial response does not mention teacher's excessive guidance. In second prompt expresses satisfaction with the guidance offered by the teacher. In third prompt, points out that child may not have needed so much help.	Weak Evidence In initial response suggests that this problem was too hard and inappropriate for this child to solve.
Evidence In initial response does not mention that the teacher did too much leading. In second prompt identifies cubes or story problem or positive reinforcement as strengths but does not talk about teacher's guidance as strength. In third response, critiques teaching for being too leading.	
Strong Evidence In initial response notes that the teacher was too leading. In third prompt criticizes teaching for being too leading.	

Figure 3. Scoring rubric for Belief 7/Segment 7.

Each belief was measured by more than one segment to give a valid measure. In the case of Belief 7, the second segment used to measure the belief included the only general (not context specific) segment: Respondents were asked, first, whether they would ever ask children to solve problems without first showing them how and, second, to explain their answer. Their explanations were assigned scores on the basis of the amount of autonomy they planned to provide for children and their rationale for doing so. Respondents who suggested that children understand more mathematics when they devise their own solution strategies were given a *Strong Evidence* score.

To determine a final score for each belief, we combined individual scores from each rubric. Because of the ordinal nature of the scores, summing scores and discussing means for each belief were inappropriate. We developed a rubric-of-rubrics system that could be applied to each belief. In this system, we accounted for the differing strengths of beliefs by having a range of scores for the rubrics and the belief scores.

Because the instrument did not employ Likert scales, traditional tests typically performed on surveys were not appropriate. We confirmed the validity and reliability of the instrument by administering it to 18 PSTs and conducting follow-up interviews. We also administered the instrument to five mathematics educators experienced at teaching and researching PSTs' beliefs. We had extensive follow-up conversations with them; they confirmed that the questions on the instrument and the rubrics used to score responses were valid measures of beliefs. When coders used the rubrics, 20% of the responses were coded by two coders, and we achieved, on average, 84% reliability on all 17 rubrics.

HOW EFFECTIVE IS THE SURVEY?

We were interested in determining whether the belief survey was sensitive enough to measure a range on each of the seven beliefs and whether the instrument would measure belief change. For 159 PSTs enrolled in the first of four mathematics courses for prospective elementary school teachers, we administered the assessment as a pretest (at the beginning of the course) and as a posttest (at the end of the course). (In our IMAP study we assigned these 159 PSTs to treatments, but presentation of those data go beyond the scope of this paper.) Pretest results indicate that most of the PSTs initially showed *no evidence* of holding each belief (see Table 1), and nearly all the PSTs (90%, 77%, and 96% for Beliefs 1, 3, and 7, respectively) fell into one of two categories, showing either no evidence or weak evidence. We found variation in scores in the pretest, showing that the instrument captured individual differences.

	No evidence	Weak evidence	Evidence	Strong evidence
B1 Pretest	60% (95)	30% (47)	9% (15)	1% (2)
B1 Posttest	19% (30)	40% (64)	25% (39)	16% (26)
B3 Pretest	64% (102)	13% (20)	18% (29)	5% (8)
B3 Posttest	28% (44)	11% (18)	26% (41)	35% (56)
B7 Pretest	71% (113)	25% (39)	4% (7)	0% (0)
B7 Posttest	40% (64)	36% (58)	19% (30)	4% (7)

Table 1. Pretest and Posttest Scores for Beliefs 1, 3, and 7 (n = 159).

Posttest results indicate that many PSTs' beliefs changed over the semester, with far fewer *No Evidence* scores and a greater number of *Strong Evidence* scores on the posttest (See Table 1). Many PSTs' responses were still coded as indicating no evidence of the beliefs. We interpret these results as indicating that our belief survey was not simply a measure of information that could be easily learned or parroted back to us by PSTs over the course of a semester in which they took a mathematics course for PSTs, as might be the case in a test of knowledge. The range of interpretations possible for each segment allowed PSTs' beliefs to emerge. Clearly, some had changed and some had not.

CONCLUSIONS

A major strength of our instrument is that it uses video clips and learning episodes to create contexts to which users respond in their own words rather than choose from one of several options. This format provides qualitative data that can be used for a variety of purposes. It also provides detailed information about the respondents' interpretations of the questions they are asked. This strength comes with a cost in terms of time required for coders to learn to use the rubrics and translate the constructed responses into quantified responses. Whether this "price" will be too high for those seeking a belief instrument is an important question left to be answered.

The mathematics of the instrument was whole number place value and rational number. We suspect that the instrument would have been different were it intended for different content, say geometry, or for a different population, say preservice secondary school teachers or practicing elementary school teachers. For example, we have given little thought to whether different approaches are needed for investigating the relationship between concepts and procedures in geometry and arithmetic. Our pilot work using the instrument with in-service teachers has been encouraging, because we found that we continued to capture a range of scores. We make no claim about the efficacy of this instrument with secondary school teachers.

Although our instrument measured change between the beginning and end of the treatment, it seemed neither “too easy” nor “too difficult”; that is, it measured neither a floor effect nor a ceiling effect. Although the high scores on the pretest were few, we did measure variation, and on the posttest we found low- and high-scoring PSTs.

Beliefs are inferred by someone who holds beliefs. The most those inferring the beliefs can do is to be clear about what those beliefs are and how those beliefs were operationalized so that others considering using the instrument can decide whether they value those beliefs and whether they agree with how those beliefs were measured. We would be presumptuous to claim that we have created an instrument to measure beliefs about mathematics and mathematics learning, so we will state only that we have created an instrument that measures seven specific beliefs about elementary school contexts. We think that we have developed a belief instrument that can effectively measure quantitative differences while still capturing the individual voice of respondents that, in the past, has been captured only through qualitative approaches.

References

- 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., & Arvola, B. (1998). Conceptualizing belief structures of preservice secondary mathematics teachers. *Journal for Research in Mathematics Education*, 29, 306–333.
- Fennema, E., Carpenter, T. P., & Loef, M. (1990). *Mathematics beliefs scales*. University of Wisconsin-Madison.
- 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.
- 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.