The Mathematics Educator
2006, Vol. 19, No. 1, 32–45

Measuring Task Posing Cycles: Mathematical Letter Writing Between Algebra Students and Preservice Teachers

Anderson Norton
Zachary Rutledge

In a secondary school mathematics teaching methods course, a research team engaged 22 preservice secondary teachers (PSTs) in designing and posing tasks to algebra students through weekly letter writing. The goal of the tasks was for PSTs to elicit responses that would indicate student engagement in the mathematical processes described by NCTM (2000) and Bloom's taxonomy (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956), as well as student engagement in the highest levels of cognitive activity described by Stein, Smith, Henningsen, and Silver (2000). This paper describes our efforts to design reliable measures that assess student engagement in those processes as a product of the evolving relationship within letter-writing pairs. Results indicate that some processes are easier to elicit and assess than others, but that the letter-writing pairs demonstrated significant growth in terms of elicited processes. Although it is impossible to disentangle student factors from teacher factors that contributed to that growth, we find value in the authenticity of assessing PSTs' tasks in terms of student engagement rather than student-independent task analysis.

Designing and posing tasks plays a central role for mathematics teaching (Krainer, 1993; NCTM, 2000). However, research indicates that preservice teachers lack ability to pose appropriately challenging mathematical tasks for students (e.g., Silver, Mamona-Downs, Leung, & Kenney, 1996). This article addresses the development of such ability by engaging preservice secondary teachers (PSTs) in posing mathematical tasks to high school algebra students through mathematical letter writing. We consider our approach an extension of the kind of letter-writing study performed by Crespo (2000; 2003). In a previous article (Rutledge & Norton, 2008), we reported results from this project related to the letter-writing interactions between PSTs and students. That article focused on comparing cognitive constructivist and socio-cultural lenses for examining the interactions. The purpose of this article is to investigate the mathematical processes that PSTs' tasks elicited from students.

Crespo (2003) engaged preservice elementary school teachers in posing mathematical tasks to fourth-grade students through letter writing. The purpose of her study was to elicit and assess students' mathematical thinking. She found the tasks preservice teachers wrote became more open-ended and cognitively complex over the weeks of letter writing. This result affirmed her key hypothesis that the preservice teachers' extended and reflective interactions with an "authentic audience" (p. 243) would provide opportunities for them to learn how to pose appropriately challenging tasks. Crespo's work informed our approach to studying the development of task-posing ability among PSTs, and we too used letter writing with PSTs to foster such development. Rather than focusing on the tasks PSTs posed, as Crespo did, we specifically examined elicited student responses as a product of the evolving relationship within letter-writing pairs.

During a secondary methods course, 22 PSTs were paired with high school algebra students; the PSTs posed tasks to their student partners and assessed the responses. As researchers, we independently examined the responses from the algebra students to make inferences about their cognitive activity. Considering this study to be an extension of Crespo's work, we introduce a method for measuring PSTs' progress in learning to design and pose individualized mathematical tasks through letter writing. We measured the effectiveness of PSTs' tasks by assessing the cognitive activities those tasks elicited from students (as indicated by student responses), and we hypothesized that such measurements would demonstrate growth over the course of letter-writing exchanges between the pairs.

We report on our design of measurements for the effectiveness of the letter-writing pairs, in addition to

Anderson Norton is an Assistant Professor in the Department of Mathematics at Virginia Tech. He teaches math courses for future secondary school teachers and conducts research on students' mathematical development.

Zachary Rutledge is completing a PhD in Mathematics Education at Indiana University. He works as an actuary in Salem, Oregon.

the results of applying that design. In particular, we relied on descriptions of cognitive activities described in three main sources: Bloom's taxonomy (Bloom et al., 1956; Kastberg, 2003), *Principles and Standards of School Mathematics* (NCTM, 2000), and a chapter on "cognitively complex tasks" by Stein, Smith, Henningsen, and Silver (2000). We chose these sources because they are common readings in the PSTs' methods courses, and they provide potential metrics for assessing the quality of tasks. We collectively modified them to form a comprehensive and complementary framework for assessing students' responses to the tasks.

In the following section we summarize the original authors' descriptions of these processes. We then describe how we operationalized the processes to assess the cognitive activities indicated by each student response. In the final two sections, we report on the reliability of our measures and the evolution of cognitive activity elicited by the PSTs' tasks over the course of 12 weeks. Findings from this study inform the following research questions: How can we reliably measure the effectiveness of the letter-writing exchanges in terms of elicited cognitive activity from the high school students? And, using the measurements we develop, in what ways do the PSTs demonstrate progress in designing and posing appropriately challenging mathematical tasks for students?

We wanted PSTs to learn to pose more engaging mathematical tasks and to assess students' thinking based on their written responses. We hypothesized that over the 12 weeks the PSTs' tasks would elicit more of NCTM's Process Standards (2000) and the four highest levels of reasoning in Bloom's taxonomy (i.e. application, analysis, synthesis, and evaluation). We also expected a general progression toward responses that indicated students were using *Procedures with Connections* and *Doing Mathematics*, moving away from responses reliant upon *Memorization* or *Procedures without Connections* (Stein et al., 2000).

Theoretical Orientation

Task Posing

Since Brown and Walter's (1990) seminal work on problem posing (read as "task posing"), many subsequent publications focused on teachers engaging students in posing problems (e.g., Gonzales, 1996; Goldenberg, 2003). Whereas these publications have implications for teacher education, they do not examine teachers' abilities to design appropriately challenging tasks for their students. Research investigating teachers' abilities to design such tasks has typically focused on student-independent attributes of the tasks, such as whether they introduce new implicit assumptions, initial conditions, or goals (Silver et al., 1996). Similarly, Prestage and Perks (2007) engaged PSTs in modifying givens and analyzing mathematical demand of tasks in order for these future teachers to develop fluency in creating ad hoc tasks in the classroom. However, Prestage and Perks noted, "the analysis of the mathematics within a task can only offer a description of potential for learning" (p. 385). Understanding the actual cognitive demand of a task depends upon the learner. "Today, there is general agreement that problem difficulty is not so much a function of various task variables, as it is of characteristics of the problem solver" (Lester & Kehle, 2003, p. 507). One such characteristic that has received insufficient attention during task posing is the students' understanding of mathematics content (NCTM, 2000, p. 5). As Crespo (2000; 2003) demonstrated, letter writing can provide a rich context for PSTs to develop task-posing ability through mathematical interactions with students and help PSTs better attend to student understanding of content.

Liljedhal, Chernoff, and Zazkis (2007) described another important component of task-posing for PSTs: "predicting the affordances that the task may access" (p. 241) as PSTs attempt to elicit particular mathematical concepts or processes from students. However, within letter writing such analyses no longer determine whether the task is 'good' because PSTs can rely on students' actual responses for making that determination. Crespo (2003) described letter writing as an opportunity for "an authentic experience in that it paralleled and simulated three important aspects of mathematics teaching practice: posing tasks, analyzing pupils' work, and responding to pupils' ideas" (p. 246). The authenticity of PST-student interactions is highly desirable because the PSTs can assess the effectiveness of their tasks without relying on the authority of a teacher educator. The benefits of this kind of authenticity might be analogous to students' experiences when they view their own mathematical reasoning as an authority, rather than relying on the text or a teacher for validation. The PSTs' problem becomes one of "witnessing the development of the activities provoked by the task, and comparing it to the ones they predicted and to the initial task" (Horoks & Robert, 2007, p. 285). This development allows PSTs to use these comparisons as they modify their initial tasks and design new tasks.

Cognitive Measures (in Theory)

In order to measure PSTs' progress in eliciting mathematical activity from students through task posing, we looked to three sources: Bloom's taxonomy (Bloom et al., 1956), the NCTM Process Standards (2000), and the levels of cognitive demand designed by Stein et al. (2000). PSTs' familiarity with these sources was important to us for the following reason: Often these sources (or others that describe a hierarchy for analyzing student thinking) are introduced to PSTs as useful ideas to adapt into their future teaching. Teacher educators should move beyond introduction of these sources and instead facilitate opportunities for PSTs to investigate ways that they prove beneficial in working with students. Therefore, we asked PSTs to assess their students' responses to tasks using these sources. This mirrors the way that we used them in this study to assess the PST's task-posing ability.

Table 1 presents the measures created for this study based on these frameworks. The short definition provides a summary of the different measures as described by the original authors. The first four measures come from Bloom's taxonomy of educational objectives (Bloom et al., 1956), the next five measures come from NCTM's Process Standards (2000), and the last four measures come from Stein et al. (2000). It is important to note that we chose to disregard the first two levels of Bloom's taxonomy, *Knowledge* and *Comprehension,* because we felt that these measures were too low-level and would likely be elicited with great frequency. On the other hand, we kept all four of Stein's levels of cognitive demand because they provide a necessary hierarchy for ranking tasks and measuring growth, as we describe in the following section.

Table 1
Measure/Process Descriptors

Measure	Short Definition
Application	Using previously learned information in new and concrete situations to solve problems
Analysis	Breaking down information into its component parts so that the relationship between parts is clear
Synthesis	Putting together elements and parts to form a whole
Evaluation	Judging the value of methods for given purposes
Communication	Expressing mathematical ideas in words to clarify and share them, so that "ideas become objects of reflection" (p. 60)
Connections	Relating mathematical ideas to each other, and to previous experiences in other domains, such as science
Problem Solving	"Engaging in a task for which the solution method is not known in advance" (p. 52), which involves the use of strategies in struggling toward a solution.
Reasoning and Proof	Making analytical arguments, including informal explanations and conjectures
Representation	The "process and product" (p. 67) of modeling mathematical ideas and information in some form, in order to organize, record, and communicate.
Memorization	Memorizing or reproducing "facts, rules, formulae, or definitions" (2000, p. 16) without any apparent connection to underlying concepts
Procedures without Connections	Using a procedure or algorithm that is implicitly or explicitly called for by the task, without any apparent connection to underlying concepts
Procedures with Connections	Using procedures to deepen understanding of underlying concepts
Doing Mathematics	Investigating complex relationships within the task, its solution, and related concepts, often involving metacognition, analysis, and problem solving

Adapted from Bloom et al. (1956), NCTM (2000), and Stein et al. (2000).

Methodology

Setting

The 22 PSTs who participated in this study were enrolled in the first of two mathematics methods courses that precede student teaching at a large midwestern university. Mrs. Rae, a local high school mathematics teacher, was interested in finding ways to challenge her students by individualizing instruction. When the PSTs' methods instructor (first author) approached her about task-posing through letter writing, Mrs. Rae agreed that such an activity would serve the educational interests of her students, as well as the PSTs. Each PST was assigned to one student from Mrs. Rae's Algebra I class, and wrote letters back and forth to her or his assigned student, once per week for seven weeks. The PSTs were given no guidelines on the type of problems to pose; instead they were instructed to focus on building students' mathematical engagement. The high school term (trimester) ended after the seventh week and students were assigned to new classes, so the PSTs began writing letters to a new group of students in the eighth week. They wrote to Mrs. Rae's Algebra II students the final five weeks of the project. Each week, the methods instructor and Mrs. Rae collected the letters and responses, respectively, and exchanged them.

In the title of this article, we use the term *cycle* to refer to the PSTs' iterative task design. After posing an initial task, we expected PSTs to use student responses to construct models of students' mathematical thinking. That is, we expected PSTs to try to "understand the way children build up their mathematical reality and the operations by means of which they try to move within that reality" (von Glasersfeld & Steffe, 1991, p. 92). Using this knowledge, the PSTs could design tasks more attuned with their students' mathematics—the students' particular mental actions and ways of applying those actions to problem-solving situations. In turn, we hypothesized that the well-designed tasks would presumably increase student engagement and cognitive activity. By focusing PSTs' attention on the cognitive activities described by NCTM, Bloom et al., and Stein et al., we hoped to provide a framework for PSTs to begin building models.

Whereas PSTs' goals for student learning often revert to mastery of procedural knowledge (Eisenhart et al., 1993), we promoted goals for conceptual learning among the PSTs through class readings and discussions. We encouraged PSTs to use open-ended tasks (i.e. tasks that invite more than one particular response) so student responses would be rich enough for PSTs to make inferences about the students' thinking. We hoped the opportunity to make inferences about the students' mathematical thinking would lead the PSTs to construct models of students' mathematics. We also encouraged the PSTs to rely on their models to imagine how students' mathematics might be reorganized in order to become more powerful, allowing the students to engage with a broader range of mathematical situations.

Data Analysis

Data consisted of PSTs' letters and students' responses. PSTs complied these documents into their notebooks, and we collected them at the end of their methods course. After removing 31 letters that were not matched with task responses, 233 tasks/response pairs remained to be analyzed.

Data analysis had four phases: (a) operationalization of our cognitive measures, (b) the raters' individual coding, (c) reconciliation of our individual coding, and (d) interpretation of the final codes. The operationalization concerns the way in which we transformed the theoretical processes given in the previous section into heuristics that allowed us to identify cognitive activity. Individual coding relied on this operationalization while continuing to inform further operationalization of the cognitive measures. As not to distort inter-rater reliability scores, in the interim we met only to discuss clarifications of the cognitive activities, without sharing notes or discussing particular responses. At the end of the letter-writing project, we computed the inter-rater reliability of our coding for the cognitive measures. Following this analysis, we reconciled our codes by arguing points of view regarding scoring differences until we reached consensus. Finally, we could interpret the reconciled codes, graphically and statistically.

Graphs of the relative frequency of each cognitive activity, as measured week-by-week, provide an indication of growth among the PST-student pairs. We use the graphs to describe patterns in elicited activity over time. Although the two different groups of students involved in letter writing (Algebra I students in the first seven weeks, and Algebra II students in the final five weeks) render a 12-week longitudinal analysis untenable, data from the two groups do provide opportunity for us to consider differences in PSTs' success in working across the groups. Finally, we performed linear regressions on aggregate results to provide a statistical analysis of progress.

Cognitive Measures (in Practice)

The operationalization occurred mostly during the individual coding phase with minor adjustments required during the reconciliation phase. That is to say, we essentially transformed the 13 measures into a system allowing a researcher or a practitioner to categorize his or her inferences of students' cognitive activity. To achieve this transformation, we began with the previously discussed definitions for the various measures and then made adjustments throughout the individual coding phase. Whenever one of the raters (authors) encountered difficulty in assessing a student response, he would approach the other to discuss the difficulty, in a general way, without referring to a particular student response. This interaction would allow the raters to decide how to resolve the difficulty and individually reassess previous ratings to ensure consistent use of the newly operationalized measure.

Table 2 describes the most fundamental changes that we made to the measures. The adjustments are the results of the following two goals: (1) to ensure that measures could be consistently applied from task to task and (2) to ensure that no two measures were redundant.

With regard to redundancy, we had to differentiate *Connections*, *Procedures with Connections*, and *Application*. We used *Connections* to refer to connections among disparate mathematical concepts; we reserved *Procedures with Connections* to describe connections between mathematical procedures and concepts; and we reserved *Application* to describe connections among mathematical concepts and other domains. With regard to our ability to consistently apply a measure, we had difficulty with *Problem Solving* and *Doing Mathematics*. As defined in Table 1, *Problem Solving* requires a struggle toward a novel solution, rather than the application of an existing procedure or concept. Lester and Kehle (2003) further characterized problem solving as "an activity requiring the individual to engage in a variety of cognitive actions" (p. 510).

Lester and Kehle (2003) described a tension between what is known and what is unknown:

> Successful problem solving involves coordinating previous experiences, knowledge, familiar representations and patterns of inference, and intuition in an effort to generate new representations and patterns of inference that resolve the tension or ambiguity (i.e., lack of meaningful representations and supporting inferential moves) that prompted the original problem solving activity. (p. 510)

Table 2
Revised Measure/Process Descriptors

Process Needing Revision	Change to Original Definition
Application	Student response must indicate use of mathematics in a domain outside of mathematics
Synthesis	Student response must indicate the production of a *new* whole, not explicitly called for in the task
Evaluation	Student response must show two or more solution methods or approaches, and compare them based on their merits in resolving the task
Memorization	Includes cases where a student seemed to look up information in a text or other resource
Doing Mathematics	Student response must indicate a new conjecture and a test of it
Connections	Student response must indicate a novel link between two or more existing mathematical concepts
Problem Solving	Student response must indicate cognitive dissonance/struggle in a novel situation leading to a resolution
Representation	Student response must indicate that the student used the representation as a cognitive aid, not simply producing a representation when explicitly asked

Note: Processes/Measures not listed were not changed during coding.

Other cognitive activities, such as representation and inference (possibly involving reasoning and proof), may support a resolution to this tension. For us to infer that a student engaged in *Problem Solving*, we needed to identify indications of this perceived tension, and we needed to infer a new construction through a coordination of cognitive actions. This meant responses labeled as involving *Problem Solving* were often labeled as involving other cognitive activities as well, such as *Analysis* and *Reasoning & Proof*.

While Stein et al.'s (2000) definition of *Doing Mathematics* provided some orientation for our work, we found it to be too vague for our purposes. So, we relied on Schifter's (1996) definition for further clarification; she defined *Doing Mathematics* as conjecturing. To infer a student had engaged in *Doing Mathematics*, we needed to infer the student had engaged in making and testing conjectures. For example, if we inferred from student work that the student designed a mathematical formula to describe a situation and then appropriately tested this formula, then we would consider this to be *Doing Mathematics*. We recognize that our restriction precludes assessment of other activities that Stein et al. (2000) would consider to be "doing mathematics," but this restriction provided a workable resolution to assessing students' written responses.

As a final and general modification of the original cognitive measures, we required that each process (other than the lowest two levels of cognitive demand) produce a novelty. For example, assessing *Synthesis* required some indication that the student had produced a *new* whole from existing constituent parts. Furthermore, we needed indication that the student generated the cognitive activity as part of their reasoning. If a PST explicitly asked for a bar graph, the students' production of it would not constitute *Representation*, because it would not indicate reasoning. Such a response would probably indicate a *Procedure without Connections*.

Results

A Letter-Writing Exchange

To illustrate the exchanges between letter-writing pairs, and to clarify the manner of our assessments, we provide the following sample exchange. A complete record of the exchange can be found in Rutledge and Norton (2008). Figure 1 shows the task posed by a PST, Ellen, in her initial letter to her student partner, Jacques. The task is similar to other introductory tasks posed by PSTs and seems to fit the kinds of tasks to which they had become accustomed from their own experiences as students. However, there is evidence (i.e. the "why" questions at the task's end) that Ellen, like fellow PSTs, attempted to engage the student in responding with more than a computational answer.

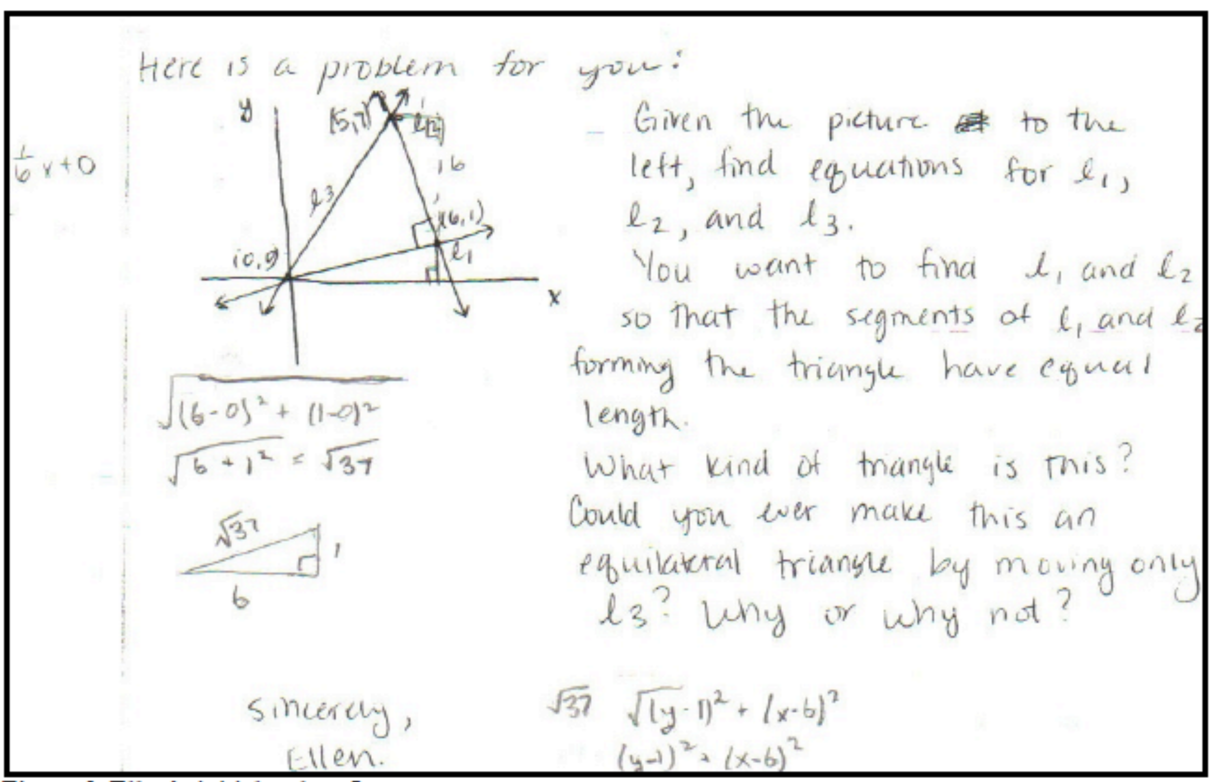

Figure 1. Ellen's initial task to Jacques

Jacques' response (Figure 2) indicates that he did not meaningfully engage in the task of finding equations for lines meeting the specified geometric conditions. However, he was able to assimilate (make sense of) the situation as one involving solutions to systems of equations. From his activity of manipulating two linear equations and their graph, we inferred that the task elicited only procedural knowledge from Jacques. It is possible that Jacques may have had a more connected understanding of the concepts underlying the procedure, but there was no clear indication from his response that allowed us to infer this. Therefore, we coded the elicited activity as *Procedures without Connections*.

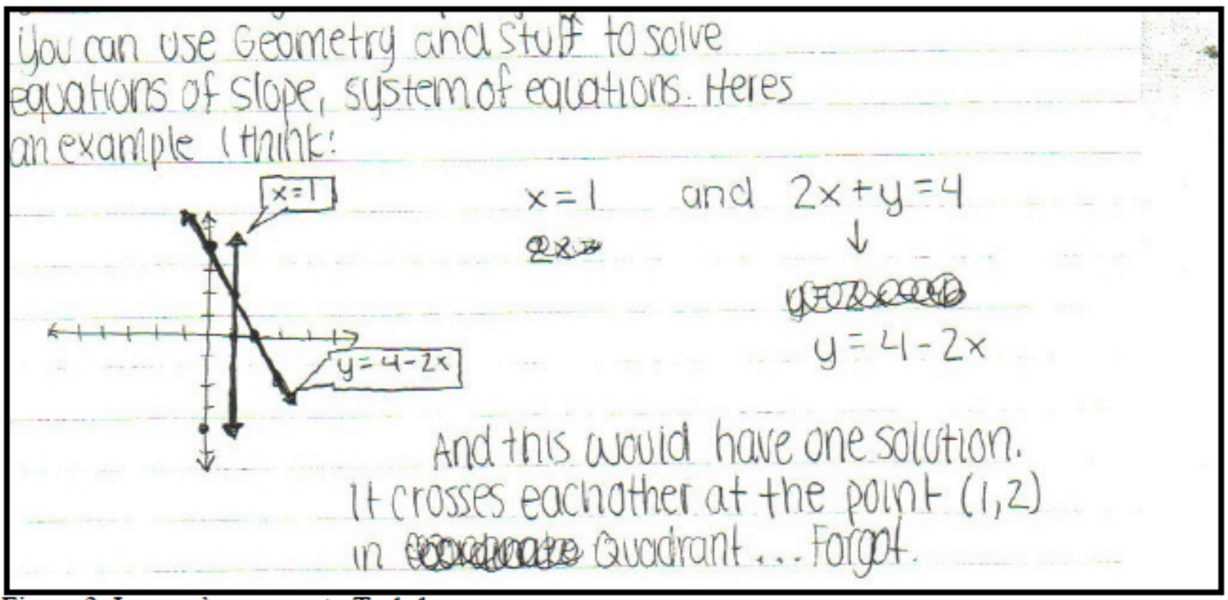

Figure 2. Jacques' response to Task 1.

Other codes assigned to Jacques' response included *Application* and *Communication*. The former was based on our inference that Jacques used existing ideas in a novel situation. He effectively applied an algebraic procedure to a new domain when he applied his knowledge of systems of equations to a situation involving finding equations of intersecting lines, When coding for *Communication*, we inferred that Jacques' written language intended to convey a mathematical idea involving the use of systems of equations to find points of intersection.

Subsequent tasks and responses indicate Ellen began to model Jacques' mathematical thinking. Using these models, she designed tasks that successfully engaged Jacques in additional cognitive activities, such as problem solving. For example, Ellen asked questions to focus Jacques' attention on the angles formed in the drawing on Figure 1. Her questions provoked Jacques to struggle through finding equations for the lines. In addition, Ellen seemed to detect an overall trend that Jacques engaged more readily with familiar procedures. She adapted to the trend and began to frame future tasks around a procedure with which she felt Jacques was likely familiar. This kind of adaptation to the student indicates that Ellen began to model the student's thinking.

Inter-Rater Reliability Results

When we finished individually coding all of the task responses for the letter-writing pairs, we met to compile the results into a spreadsheet, for this process allowed us to measure inter-rater reliability and reconcile discrepancies. To understand inter-rater reliability, we considered three measures as shown in Tables 4 and 5. These measures were Cohen's Kappa, Percent Agreement, and Effective Percent Agreement. As Table 4 indicates, percent agreement was high on all four measures from Bloom's taxonomy, but this result is because of the rarity of either rater identifying the measures. Effective percent agreement provides further confirmation of this outcome by considering agreement among those items positively identified by at least one of the raters.

Moving from left to right, the Kappa scores in Table 3 show decreasing inter-rater reliability as we progress to higher levels of Bloom's taxonomy. Sim and Wright (2005) cite Landis and Koch who suggest the following delineations for interpreting Kappa scores: less than or equal to 0 poor, 0.01-0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80

substantial, and 0.81-1 almost perfect. With this in mind, we see *Application* has a substantial agreement, *Analysis* has moderate agreement, *Synthesis* has slight agreement, and *Evaluation* did not have agreement distinguishable from random.

The high percent agreement in combination with a low effective percent agreement for *Synthesis* and *Evaluation* highlight the fact the raters rarely identified these two constructs. In the few times one rater identified such a construct when the other did not, this resulted in a low Kappa score. For further support of this assessment, we note that the 95% confidence interval for these two measures includes 0; thus, there is no support for inter-rater reliability with these two measures.

Table 4 displays the data for the NCTM Process Standards. In a similar analysis, *Communication* and *Problem Solving* have *moderate* agreement. *Representation* is on the border between *fair* and *moderate*; *Reasoning & Proof* is on the border between *slight* and *fair*; and *Connections* has a poor inter-rater reliability. *Connections* and *Reasoning & Proof* were quite rare, hence the high levels of percent agreement and the lower level of Kappa, as was the case with *Synthesis* and *Evaluation*. In fact, the 95% confidence interval for each of these measures includes 0, indicating no reliability on these measures.

Table 5 displays the raters' responses for the levels of cognitive demand described by Stein et al. (2000). The Kappa was .55—described as moderately reliable—with a 95% confidence interval of .47 to .64 (note that we report only one Kappa because we could choose only one categorization for each task response). We see that although 164 of the 233 items are on the main diagonal (showing agreement), there is definite spread away from the diagonal as well. We determined this was partly attributable to a shift in how conservatively the raters interpreted student responses. Specifically, one rater tended to identify items as eliciting lower cognitive demand than the other rater. This is seen in the total column and row, where one rater identified 178 items as either *Memorization* or *Procedures without Connections* and only 45 items as *Procedures with Connections*. Alternately, the other rater found 157 items to be either *Memorization* or *Procedures without Connections* and 70 items to be *Procedures with Connections*.

Table 3
Inter-rater Reliability for Bloom's Taxonomy

	Application	Analysis	Synthesis	Evaluation
Kappa	0.674	0.443	0.040	-0.013
Percent Agreement	86.7%	76.8%	91.0%	97.4%
Effective Percent Agreement	62.2%	43.2%	4.5%	0.0%
95% Confidence Interval (Kappa)	(0.57, 0.78)	(0.31, 0.57)	(-0.35, 0.43)	(-0.81, 0.79)

Table 4
Inter-rater Reliability for NCTM Process Standards

	Communication	Connections	Representation	Reasoning & Proof	Problem Solving
Kappa	0.567	0.051	0.388	0.200	0.555
Percent Agreement	79.4%	85.0%	89.3%	89.7%	88.0%
Effective Percent Agreement	58.3%	5.4%	28.6%	14.3%	45.1%
95% Confidence Interval (Kappa)	(0.46, 0.68)	(-0.24, 0.34)	(0.16, 0.61)	(-0.1, 0.5)	(0.4, 0.71)

Table 5
Raters' Responses for the Stein et al. Measures

	M	P	C	D	Total
M	52	14	3	0	69
P	0	80	29	0	109
C	0	11	30	4	45
D	0	0	8	2	10
Total	52	105	70	6	233

Note. M = "Memorization," P = "Procedures without Connection," C = "Procedures with Connections," and D = "Doing Mathematics"

To summarize, *Evaluation* was the measure with the weakest reliability, for it had a negative Kappa associated with it. Although having a positive Kappa, *Synthesis, Connections,* and *Reasoning & Proof* had confidence intervals that contained 0; this result indicates we cannot be certain whether it was above 0 randomly. It is also important to note these measures were some of the least-often identified. Conversely, the most reliable measures were *Communication*, *Problem Solving*, and *Application*. In addition, our assessments of levels of cognitive demand were reliable to a similar degree. Again, this conclusion is supported by our design in that these measures were more commonly identified in the coding.

Elicited-Response Results

After measuring inter-rater reliability, we reconciled our scores by arguing for or against each discrepant score. For example, the second author assessed Task 1 (Figure 1 and Figure 2) as having elicited *Connections*. However, the first author successfully argued that the evidence was stronger for connections to concrete situations, and according to our operationalization that should be coded as *Application*. We report on the reconciled scores in Figures 3, 4, and 5. Each of those figures illustrates the percentages of responses that satisfied our negotiated measures over the course of the twelve weeks. We excluded missing responses from all calculations. Although we included letters from week 1, we note many of the introductory letters did not include tasks, presumably because the PSTs were becoming familiar with the students and the format of the activity. We also note the first seven letters were written between PSTs and Algebra I students, whereas the final five letters were written between PSTs and Algebra II students. Once again, the letters written in week 8 were introductory letters, though these included many more tasks. In Figures 3, 4, and 5, the dark vertical line

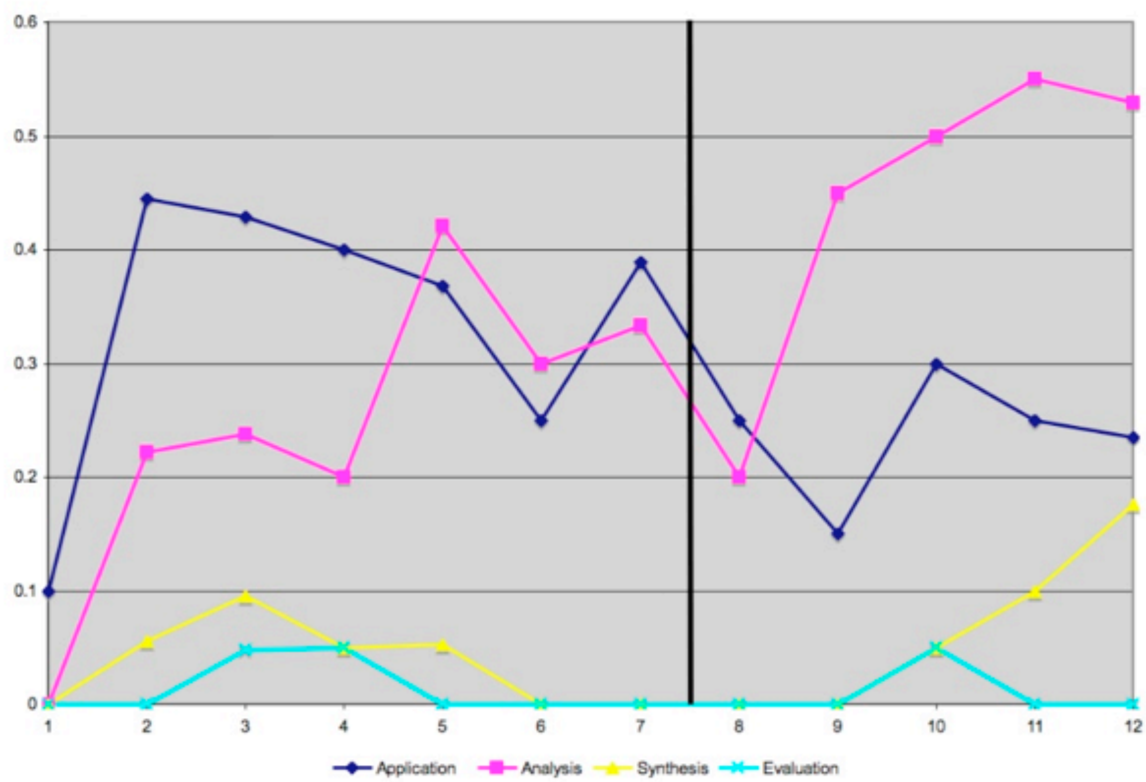

Figure 3. Results of Measures Using Bloom's Taxonomy

between weeks 7 and 8 marks the transition from Algebra I letters to Algebra II letters.

Ignoring the introductory letters from week 1, the general trends illustrated in Figure 3 indicate the frequency of PSTs eliciting *Application* from students decreased over the duration of letter writing, even across correspondences with Algebra I and Algebra II students. *Analysis* was elicited more frequently, with a pronounced spike among correspondence between PSTs and Algebra II students in the final weeks. As previously mentioned in the reliability results, we found both *Synthesis* and *Evaluation* were rarely elicited in correspondence with either group of students.

These patterns indicate the levels of cognition described by Bloom's taxonomy—at least in our operationalization of them—were heavily dependent on the PSTs and the tasks they posed. Interestingly, these patterns show little apparent dependence on the groups of students (i.e. Algebra I and Algebra II). This outcome may be because many of the posed tasks inherently required application and analysis to resolve

them, with PSTs gaining a greater appreciation for students' use of analysis over the course of the semester. *Application* tasks tended to describe new situations where the PSTs inferred, often correctly, that the students could use existing knowledge. For example, we characterized Jacques' response in Figure 2 as indicating an *Application*, for he applied his knowledge about solving linear equations to a concrete situation that required finding the point of intersection of two lines. *Analysis* tasks often involved equations whose components needed to be examined. For example, in a subsequent exchange with Ellen, Jacques broke down the triangle (Figure 1) into three lines and correctly identified the sign of the slopes of these lines. Using this knowledge, he attempted to formulate the equations of these lines. It seems that either PSTs were less familiar with the kinds of tasks that might elicit *Synthesis* and *Evaluation*, or students did not readily engage in such activity.

In Figure 4, we begin to see some differences between the elicited responses of the two groups of students. Whereas *Connections*, *Representation*, and *Reasoning & Proof* were rarely elicited from either group of students, there is a pronounced increase in *Problem Solving* among correspondence with the Algebra II students as compared to the Algebra I students. *Communication* also increased among correspondences with Algebra II students, but seemed to be elicited in a pattern that was similar across the two groups.

We see mathematical communication from both groups of students increased to a peak during the

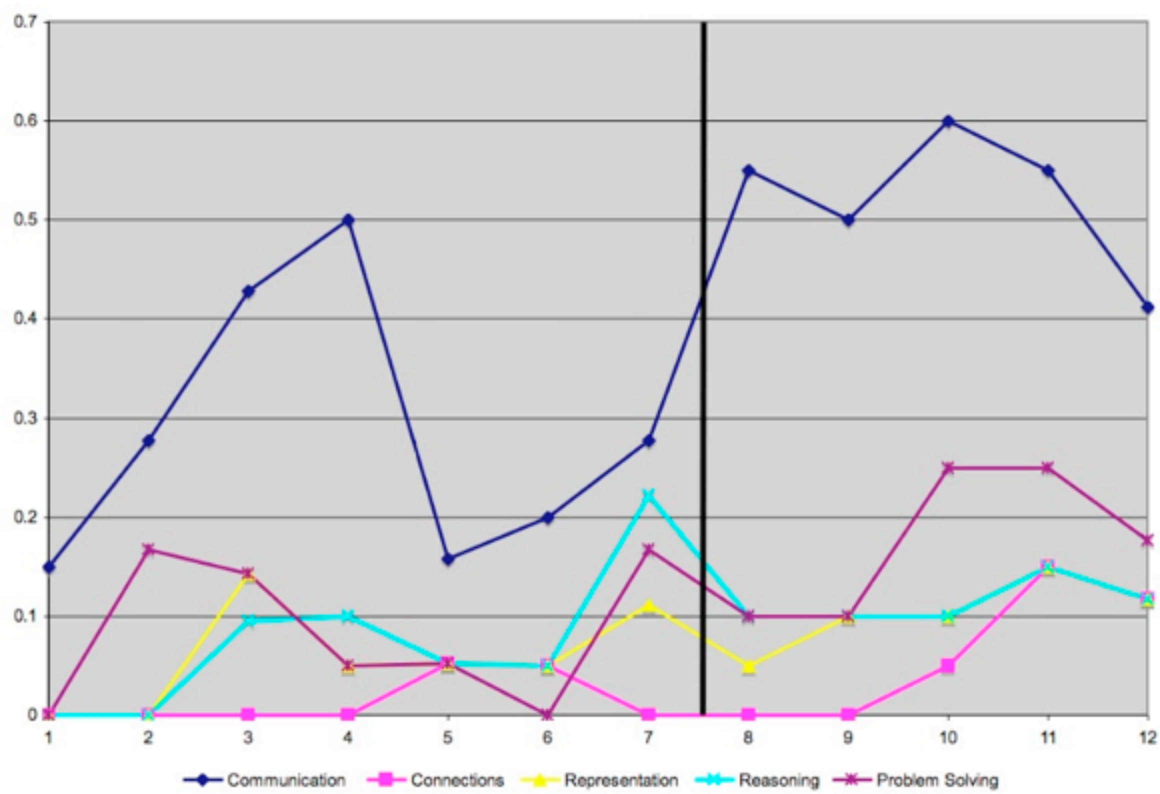

Figure 4. Results of Measures Using NCTM Process Standards

middle weeks, and then decreased toward the end of the correspondence between letter writing pairs. This trend may be due to the PSTs' initial interest in the students' thinking, which was replaced by more goal-directed tasks, once the PSTs determined a particular trajectory along which to direct the students. We found *Communication* dropped among both groups of students after their fourth week. This could be due to a lack of enthusiasm among the students once the novelty of letter writing had faded. In fact, Mrs. Rae noticed the Algebra I students began to tire of writing responses and wrote less in later weeks.

Figure 5 illustrates a general trend away from tasks eliciting *Memorization*. It seems the PSTs used students' recall of facts in order to gauge where the students were developmentally, both at the beginning and end of their correspondence with the students.

Procedures without Connections dominated the elicited responses from students, whereas *Procedures with Connections* seemed to play a significantly lesser role. It is also interesting to note that the few instances identified as *Doing Mathematics* occurred among correspondence with Algebra II students. Along with the previous observation about *Problem Solving* (namely, that problem solving was elicited much more with Algebra II students), these results lead us to one of two conclusions: (1) the PSTs held higher expectations for Algebra II students (in terms of cognitive activity, and not just content) and were, therefore, more inclined to challenge them with higher-level tasks, or (2) the Algebra II students were better prepared (either from previous learning or accepted social norms) to engage in these higher levels of cognitive activity.

A Statistical Analysis

In addition to considering the measures individually, we performed a linear regression on aggregate results over time. The first column in Table 6 lists the average number of processes elicited week-by-week, among the five NCTM Process Standards and the four highest levels of Bloom's taxonomy. For example, the PSTs elicited, on average, two of the nine processes during Week 10. The second column in Table 6 lists the average ranking of the levels of cognitive demand elicited week-by-week. We ranked *Memorization* as 0, *Procedures without Connections* as 1, *Procedures with Connections* as 2, and *Doing Mathematics* as 3. If we accept this simple form of ranking, the student responses from Week 7 indicated, on average, as a 1, *Procedures without Connections*.

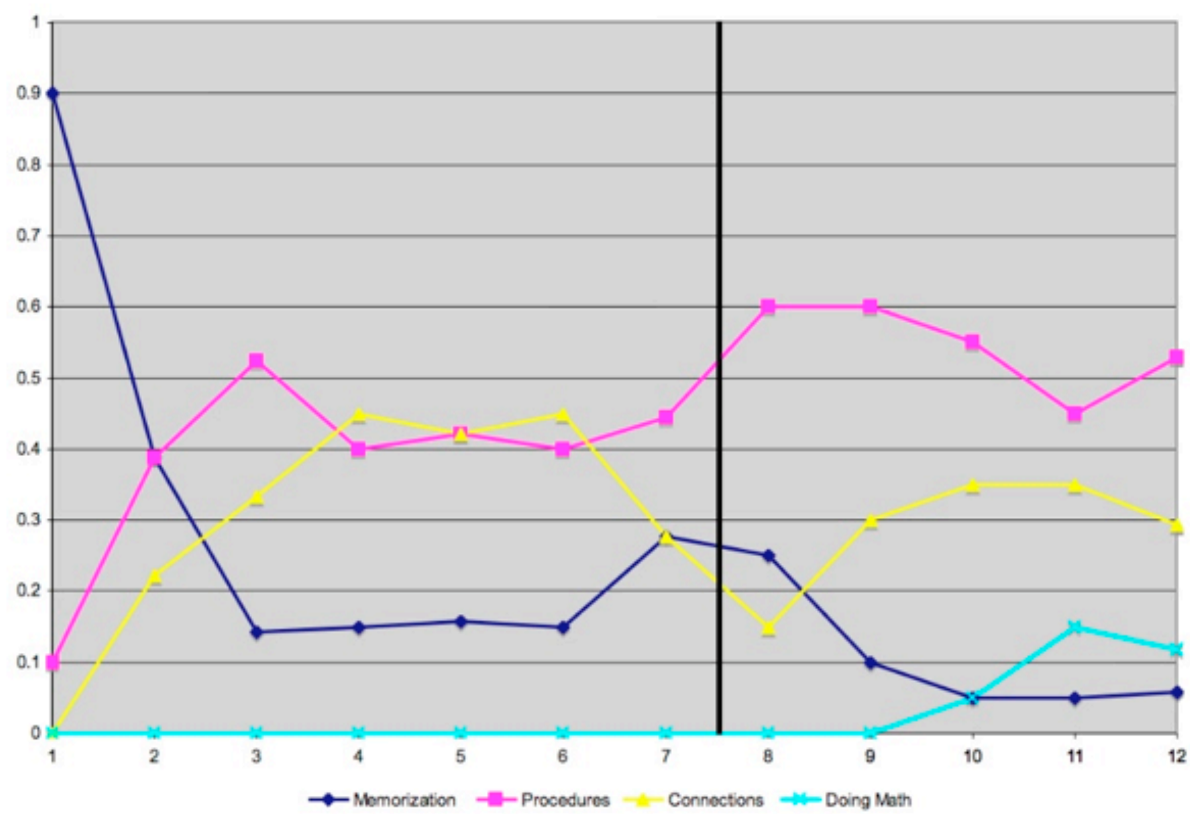

Figure 5. Results from Measures Using Stein's Categorization

Table 6
Aggregate Results for Cognitive Measures

Week	Average	
	Number of Processes	Category Ranking
1	0.25	0.10
2	1.16	0.83
3	1.62	1.19
4	1.40	1.30
5	1.21	1.26
6	0.90	1.30
7	1.50	1.00
8	1.25	0.90
9	1.40	1.20
10	2.00	1.40
11	2.15	1.60
12	1.88	1.47

Table 7 reports the slopes and *r*-squared values for each column over each of the following time periods: the first seven weeks (interactions with Algebra I students), the final five weeks (interactions with Algebra II students), and the entire 12 weeks (across the two groups of students). In addition, Table 7 includes the corresponding *p*-values to indicate whether the slopes are statistically significant. We calculated these values using rank coefficients. The slopes provide indications of the group's growth from week to week to the degree that the *r*-squared values approach 1 and *p*-values approach 0.05. It is interesting to note that the slope, *r*-squared value, and *p*-value for the final five weeks of letter writing suggest considerable growth in the level of mathematical engagement during interactions between PSTs and Algebra II students.

Discussion of Findings and Implications

Having operationalized measures of cognitive activity and having applied them to a cohort of letter-writing pairs, we are now prepared to evaluate the measurements. We intend to improve the measurements in terms of their reliability and their value as assessments of professional growth. First, we recognize areas of weakness in reliable uses of the measures, as well as areas of weakness in elicited responses. These areas coincide because cognitive activities that were least assessed were assessed least reliably; they include *Synthesis*, *Evaluation*, *Connections*, *Reasoning & Proof*, and *Representation*.

Reliability of Measures as Operationalized

When we reconciled our independent assessments of task responses, common themes emerged concerning the least assessed cognitive activities. Some of these involved highly subjective judgments, such as the novelty of the activity for the student and the student's familiarity with particular concepts and procedures. This subjectivity highlights the need for us to make our assessments based on inferences about the student's mathematical activity, just as we asked the PSTs to design their tasks based on such inferences. For example, one student assimilated information from a story problem in order to produce a simple linear equation. During coding, this response would typically be evaluated as *Synthesis*; however, one of the raters inferred that the student was so familiar with the mathematical material that her actions indicated a procedural exercise.

For our reconciliations of all measures, we agreed each cognitive activity needed to produce a mathematical novelty, such as a tabular representation a student produced to organize data in resolving the task. If the PST were to request the table, then the student's production of it would not be considered novel. Therefore, this response would not be labeled as a *Representation*. Likewise, we decided *Connections* should be used to refer to a novel connection between two mathematical concepts, such as a connection a student might make between his or her concepts of *function* and *reflection* in resolving a task involving transformational geometry. The need to make inferences about the novelty of students' activities introduced ambiguity in assessing student responses,

Table 7
Slopes and R-Squared Values for Linear Regressions

Group	Processes			Categories		
	Slope	r^2	p	Slope	r^2	p
Algebra I	0.10	0.22	0.090	0.04	0.18	0.078
Algebra II	0.20	0.66	0.081	0.15	0.89	0.036
All	0.11	0.74	0.166	0.04	0.39	0.099

particularly because we assessed responses week-by-week without considering the history of each student's responses.

We also realized we introduced some reliability issues through our selection of cognitive activities. Whereas we were pleased with the diversity of measures offered by the three categorizations (Bloom's taxonomy, NCTM's Process Standards, and Stein's levels of cognitive demand), these are not mutually exclusive. Despite our efforts to operationalize the measures in a way that would reduce overlap, we realized, for example, *Connections* would always implicate *Procedures with Connections* or *Doing Mathematics*. Additionally, we recognize that *Doing Mathematics* would always implicate *Problem Solving*, and *Reasoning & Proof* would always implicate *Communication*. Recognizing such implications might reduce ambiguity and increase reliability by eliminating some of the perceived need for raters to choose one measure over another. Finally, because frequently elicited cognitive activities were measured reliably, we conclude that supporting PSTs' attempts to elicit all cognitive activities can increase the reliability of each measurement. This support would also promote our goals for PSTs to design more engaging tasks.

Eliciting Cognitive Activity

We originally hypothesized that PSTs' tasks would elicit more cognitive processes over time, and we anticipated a general progression toward the highest levels of cognitive demand. Such findings would indicate growth in the evolving problem-posing/problem-solving relationships between PSTs and students. Our hypothesis is confirmed to the degree that *r*-squared values indicate the positive slopes reported in Table 7. Those values indicate that the relationships were particularly productive between PSTs and Algebra II students. There are many reasons students' content level might have influenced the relationship, and we cannot discern the main contributors. Possible contributors include the following: (1) PSTs wrote to the Algebra II students second and for a shorter duration so the students' remained motivated throughout the project; (2) greater content knowledge of Algebra II students contributed to greater process knowledge as well by allowing the students to engage in more problem solving or make more connections; (3) PSTs were more familiar with the content knowledge of Algebra II students so the PSTs were better prepared to design more challenging tasks; (4) social norms in the two classes differed and affected students' levels of engagement. In any case, our findings do indicate that PSTs—as a whole and over the course of the entire twelve weeks—became more successful in eliciting cognitive activity through their letter writing relationships.

Our findings also indicate which cognitive activities seem most difficult to elicit through letter writing, and we have suggested classroom social norms play a role in students' reluctance to engage in some of those activities, such as *Problem Solving* and *Doing Mathematics*. However, we also found that PSTs were able to engage students in some cognitive activities, such as *Application* and *Communication*, which affirms, "prospective teachers have some personal capacity for mathematical problem posing" (Silver et al., 1996, p. 293). Moreover, PSTs demonstrated increased proficiency at engaging their student partners in additional higher-level cognitive activities, such as *Analysis*.

Silver et al. found, "the frequency of inadequately stated problems is quite disappointing" (1996, p. 305). Although, like Silver et al., our expectations for our PSTs' task design were not met, we found students accepted nearly all of the tasks as personally meaningful and engaged in some kind of mathematical activity as a result. The disparity of this finding with that of Silver et al. (1996) might be attributed to our disparate approaches in studying problem posing. Most notably, the PSTs in our study designed tasks with particular students in mind and used student responses to assess the effectiveness of those tasks and to model students' thinking. We believe such experiences are essential to making methods courses personally meaningful to future teachers.

References

Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. New York: Longmans.

Brown, S. I., & Walter, M. I. (1990). *The Art of Problem Posing*. Hillsdale, NJ: Erlbaum.

Crespo, S. (2000). Seeing more than right and wrong answers: Prospective teachers' interpretations of students' mathematical work. *Journal of Mathematics Teacher Education*, *3*, 155–181.

Crespo, S. (2003). Learning to pose mathematical problems: Exploring changes in preservice teachers' practices. *Educational Studies in Mathematics*, *52*, 243–270.

Eisenhart, M., Borko, H., Underhill, R., Brown, C., Jones, D., & Agard, P. (1993). Conceptual knowledge falls through the cracks: Complexities of learning to teach mathematics for understanding. *Journal for Research in Mathematics Education*, *24*, 8–40.

Goldenberg, E. P. (2003). Problem posing as a tool for teaching mathematics. In H. L. Schoen (Ed.), *Teaching mathematics through problem solving: Grades 6-12* (pp. 69–84). Reston, VA: National Council of Teachers of Mathematics.

Gonzales, N. A. (1996). A blueprint for problem posing. *School Science and Mathematics*, *98*, 448–456.

Horoks, J., & Robert, A. (2007). Tasks designed to highlight task-activity relationships. *Journal of Mathematics Teacher Education*, *10*, 279–287.

Kastberg, S. (2003). Using Bloom's taxonomy as a framework for classroom assessment. *The Mathematics Teacher*, *96*, 402–405.

Krainer, K. (1993). Powerful tasks: A contribution to a high level of acting and reflecting in mathematics instruction. *Educational Studies in Mathematics*, *24*, 65–93.

Lester, F. K., & Kehle, P. E. (2003). From problem solving to modeling: The evolution of thinking about research on complex mathematical activity. In R. Lesh & H. M. Doerr (Eds.), *Beyond constructivism* (pp. 501–518). Mahwah, NJ: Lawrence Erlbaum.

Liljedahl, P., Chernoff, E., & Zazkis, R. (2007). Interweaving mathematics and pedagogy in task design: A tale of one task. *Journal of Mathematics Teacher Education, 10*, 239–249.

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

Prestage, S., & Perks, P. (2007). Developing teacher knowledge using a tool for creating tasks for the classroom. *Journal of Mathematics Teacher Education*, *10*, 381–390.

Rutledge, Z., & Norton, A. (2008). Preservice teachers' mathematical task posing: An opportunity for coordination of perspectives. *The Mathematics Educator, 18(1), 31–40*.

Schifter, D. (1996). A constructivist perspective on teaching and learning mathematics. In C. T. Fosnot (Ed.), *Constructivism: Theory, perspectives, and practice* (pp. 73–80). New York,: Teachers College Press.

Silver, E. A., Mamona-Downs, J., Leung, S. S., & Kenney, P. A. (1996). Posing mathematical problems: An exploratory study. *Journal for Research in Mathematics Education*, *27*, 293–309.

Sim, J., & Wright, C. C. (2005). The Kappa Statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*, 257–268.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development*. New York: Teachers College Press.

von Glasersfeld, E., & Steffe, L. P. (1991). Conceptual models in educational research and practice. *The Journal of Educational Thought*, *25*(2), 9–103.