DRAFT

Measuring Teacher Knowledge in Mathematics Professional Development Using Embedded Assessments

Daniel J. Heck (Horizon Research, Inc.)

Paper presented at the Annual Meeting of the American Educational Research Association

April, 2003

Chicago, Illinois

Author Contact Information:

dheck@horizon-research.com

919-489-1725

Horizon Research, Inc.
326 Cloister Ct
Chapel Hill, NC 27514

The work presented in this paper has been supported by the Core Evaluation of the Local Systemic Change through Teacher Enhancement (National Science Foundation REC-9912485) and conducted in cooperation with the VideoCases for Mathematics Professional Development (National Science Foundation ESI- 9731339). The approach taken and the opinions expressed in this paper are solely the author's and do not reflect the positions or opinions of the funders or collaborators.

Measuring Teacher Knowledge in Mathematics Professional Development Using Embedded Assessments

Introduction

The purpose of this paper is to present measurement and evaluation data from a pilot study of assessing teacher knowledge in mathematics professional development. The context of the study is the VideoCases for Mathematics Professional Development (VCMPD) program. VCMPD (NSF ESI- 9731339) is a curriculum for the professional development of teachers of mathematics in the middle and early high school grades. The teacher knowledge measured in the study is keyed to the learning goals of VCMPD. The instruments for measuring teacher knowledge are open-ended tasks embedded within the VCMPD program. VCMPD contains several modules of professional development. The module in which this study was conducted focuses on the mathematics and the teaching of linear relationships.

Two questions are addressed in the paper:

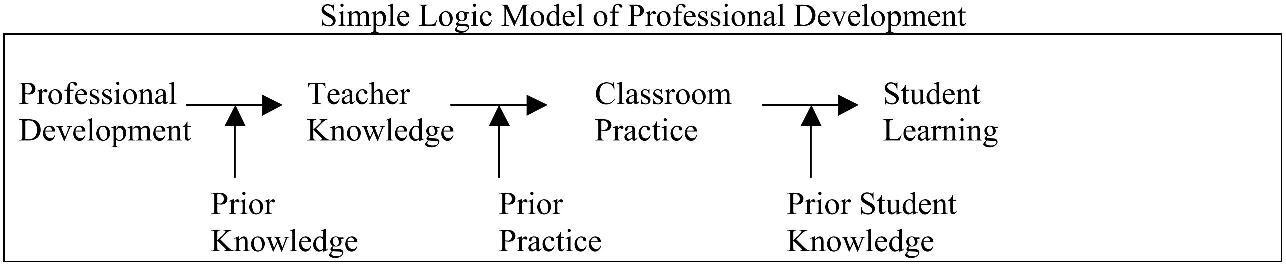
(1) How can embedded, open-ended tasks be used to assess teacher knowledge in professional development?

(2) Based on embedded, open-ended measures of teacher knowledge, what effects of the VCMPD program can be identified?

Conceptual Background

The premise of providing professional development for teachers to improve education is represented is the logic model in Exhibit 1. The research-based link between teacher professional development and changes in classroom teaching and student learning is tenuous at best (Kennedy, 1998; Wilson & Berne, 1999). Even where links to either practice or student learning have been established, little is known about what teachers actually learn in professional development that may lie behind changes in teaching (Wilson & Berne, 1999).

Exhibit 1
Simple Logic Model of Professional Development



The design of the VCMPD program flows from a specific set of learning goals for teachers:

(1) Improving teachers' disciplinary content knowledge as it relates to teaching;
(2) Improving teachers' ability/propensity to forecast student thinking about mathematics tasks;

(3) Improving teachers' ability/propensity to identify and categorize student thinking in ways useful to making decisions about teaching mathematics;
(4) Improving teachers' ability/propensity to select mathematics tasks suited to teaching particular mathematics concepts; and
(5) Improving teachers' ability/propensity to apply a repertoire of appropriate pedagogical strategies to teaching particular mathematics concepts.

VCMPD's learning goals and design are directly related to the logic model of professional development in Figure 1. A clear tenet of the learning goals and the design of VCMPD is that teachers' experiences in professional development must be closely connected to their responsibilities in the classroom. In this respect, VCMPD is a good example of the emerging breed of teacher professional development called "practice-based." The central notion of practice-based professional development is that teachers' learning experiences are more likely to result in substantive and positive changes in classroom practices when they are situated in the work that teachers do in classrooms, and focus on the most important aspects of that work (e.g., deep understanding of content, understanding student thinking about content, selecting effective pedagogical strategies to aid learners in achieving important learning goals). In VCMPD this connection is largely addressed through the central place of "classroom artifacts" in the professional development. Teachers in the program spend a considerable amount of time engaging with videos and other samples of teacher and student work. In this way, the program situates the professional development experience in the work of classroom teachers, making a connection between professional development and prior knowledge toward the aim of improving teachers' resulting knowledge. Moreover, the program devotes considerable time to having teachers consider implications for their own practice, making a connection between new knowledge, prior practice, and future practice.

All of VCMPD's goals are linked to enhancing teachers' knowledge of mathematics, and specifically the mathematical knowledge most pertinent to the teaching of mathematics. That is, the kinds of mathematical knowledge targeted, and the intention to improve teachers' propensity to apply their mathematical knowledge in a variety of ways, are all directly related to the work of teaching. In this way, the goals of VCMPD address the important work not only of deepening teachers' knowledge of mathematics, but also of enhancing teachers' ability and propensity to use that knowledge to understand students' mathematical ideas, and to plan instruction that addresses key ideas in developing students' conceptual understanding (Ball & Hyman, 2000; Carpenter, Fennema, Peterson, & Carey, 1988; Fennema & Franke, 1992; Franke, Carpenter, Levi, & Fennema, 2000; Ma, 1999; Putnam & Borko, 2000).

Context of the Study

The VCMPD program has been in development since 1997. In 2001-02, the first full module was pilot tested. The module focuses on teaching the mathematics of linear relationships. It consists of 8 sessions of approximately 3 hours each. Each session includes time devoted to situating the work of the session with respect to the work of teaching mathematics, doing mathematics, viewing and analyzing videocases of classroom practice, and linking the experience and learning to teachers' practice.

During the pilot testing of the module, the developers sought a variety of ways to understand what teachers were learning from their experiences and how that learning was translated to classroom practice. The developers and evaluation consultants from Horizon Research, Inc.

fashioned the idea of using tasks already occurring in the VCMPD program as embedded assessments to measure teacher knowledge in professional development. The pilot testing of the module was also used to pilot test the embedded assessments.

In the fall of 2002, a revised version of the VCMPD module on linear relationships was piloted. During the second pilot, the embedded assessments were employed to measure teacher knowledge in the professional development program.

Instruments

The measures of teacher knowledge used in this study occur in the context of teacher professional development. The measures are called "embedded assessments" because all of the data gathered are a part of the work teachers do in the VCMPD program. The measures are intended to gather some information about teachers' prior knowledge and resulting knowledge related to key goals targeted by a specific professional development program. Many of the tasks also ask teachers to apply their knowledge to situations of classroom practice.

The tasks were developed by the developers of the VCMPD program in collaboration with evaluation consultants from Horizon Research, Inc. Several of the tasks that existed in the materials were adapted in order to serve as embedded assessments. Some other tasks were added by developing them first as assessments, but then incorporating them into the professional development program

One objective of using embedded assessments was to minimize potential negative influences on the professional development culture that might be introduced through "testing" teachers. The instruments are completed on carbonless copy paper, with teachers keeping the original and returning the copy to facilitators. In this way, teachers' work on the assessments became a resource to use in discussions and reflections and a record of teachers' own work in the program. Also, the copies of the instruments submitted to facilitators were anonymous.

Five instruments were used in the study. Three of the instruments are Mathematics Tasks. On these measures, teachers are presented with a mathematics problem similar to one they might use with their students. The Mathematics Tasks consist of teachers solving the problem, forecasting ways that students might think about the problem differently, and identifying the mathematical ideas that the problem might be good for teaching. The Mathematics Tasks were administered in sessions 1, 6, and 8 of the VCMPD program.

Another instrument used in the study was a VideoCase Analysis. For this measure, teachers watched a segment of classroom practice in which a teacher and students are working on one of the mathematics tasks. Teachers are asked to identify an interchange in the VideoCase that they found mathematically interesting or important. In their written response, teachers describe why they found the interchange mathematically important, and what the teacher and/or students might be thinking about the mathematics of the task during that interchange. The same VideoCase Analysis instrument was administered in sessions 1 and 8.

Finally, a Linking to Practice instrument was used in the study. For this measure, teachers are given a mathematics problem and two learning objectives for using the problem in instruction. Four samples of student work that are on the problem are provided as representative of the work of a class of 25. Teachers are asked to provide their assessment of the class's current level of understanding of the learning objectives based on the samples of student work, and to describe a plan for a subsequent lesson targeting the same learning objectives based on their assessment of the student work. This instrument was administered during session 7.

Using response to instruments from the pilot testing of the embedded assessments, the developers and author identified the type and nature of evidence provided in responses related to the learning goals of the VCMPD program. From this process, a set of criteria for scoring response to each task in relation to each relevant learning goal were initially created. In the process of training raters to score responses, the developers and author made slight modifications to the criteria in order to clarify meanings.

Scoring of the instruments was conducted using rubrics keyed to learning goals of the VCMPD program. Three rubrics were used for response to the Mathematics Tasks instruments; 2 for responses to the Linking to Practice instrument; and 1 for responses to the VideoCase Analysis instrument. Each rubric consisted of 5 to 10 criteria identifying specific aspects of the response related to teachers' ability/propensity to provide evidence of that aspect of knowledge in their response. Some of the criteria addressed aspects of the response that were specifically prompted on the instrument (interpreted mainly as ability to provide evidence) and others addressed aspects of the response that were not specifically prompted, but are aligned with the goals of the VCMPD program (interpreted mainly as propensity to provide evidence). Responses were scored on each criterion using a three-point scale. For each criterion, the response was rated:

2 = Yes = The response provides clear evidence to meet this criterion.
1 = Somehwhat = The response provides some evidence to meet this criterion.
0 = No = The response provides no evidence to meet this criterion.

Sample

Responses were gathered from teachers of mathematics participating in VCMPD in four sites and teachers serving as a control group in one site. A total of 53 teachers—46 participating teachers and 7 control group teachers—provided responses on one or more instruments.

Participating teachers averaged 13.6 years of experience in teaching. Nearly all taught grades between 6 and 10, with the most teaching at grades 7 and 9. Approximately 55% of participating teachers held a degree with a major or minor in mathematics; and approximately 82% held a degree with a major or minor in education. Participating teachers received a stipend for their time and effort in VCMPD.

Because the participating teachers all volunteered to participate in the VCMPD program, a control group was recruited to account for the "volunteer" factor. Teachers who had indicated an interest in participating in VCMPD in the pilot or in future sessions were asked to serve in the control group. Control group teachers had an average of 11.3 years of teaching experience. They all taught in grades 7 -10. In this group, approximately 71% held a degree with a major or minor in mathematics; and the same percentage held a degree with a major or minor in education. Teachers in the control group received the same stipend as the participating teachers for their contribution.

All responses of teachers were scored and used to analyze interrater reliability data. Responses of teachers completing all instruments of each type were used to analyze evidence of program effects on the learning goals. The number of responses included for each instrument for this analysis is reported in Table 1.

Table 1
Responses by Instrument

Instrument	Number of Responses
Mathematics Task 1	53
Mathematics Task 6	24
Mathematics Task 8	16
VideoCase Analysis 1	49
VideoCase Analysis 8	36
Linking to Practice 7	35

Design

Administration of instruments

The embedded assessment instruments were administered by the VCMPD pilot facilitators. These individuals were professional development providers in the pilot sites, and were trained by the developers in facilitation of the VCMPD module. The author also worked with the facilitators on administration of the embedded assessments to assure that standard procedures were followed. Each instrument was administered to teachers at a specified time in the VCMPD module. Teachers were issued anonymous identification labels to use on the instruments. Teachers were instructed to work individually on the instrument for a specific amount of time. Copies of the teacher responses were collected prior to any discussion of the tasks.

Control group teachers met with the facilitator in their site two times, with the first and last meeting occurring about 8 weeks apart in order to match the intended time between the first and last sessions of the VCMPD module. These teachers completed Mathematics Task 1 and VideoCase Analysis 1 during the first meeting and completed Mathematics Task 6, Mathematics Task 8, Linking to Practice 7, and VideoCase Analysis 8 during the last meeting. In the interest of receiving responses to all instruments from all control group teachers, some were allowed to complete instruments outside of meeting time, but were instructed to adhere to the specified time and other guidelines for completing the instruments.

Scoring of responses

Four mathematics education consultants served as raters. Each consultant had a background in K-12 teaching and in mathematics education evaluation or research. The raters were trained to use the rubrics during a one-and-a-half day meeting conducted by the author. The developers of the VCMPD program also attended and contributed to the training.

Following the meeting, raters were sent one of two responses to each instrument to score, so that each rater scored a response from each task, and each response was scored twice independently. For these responses, the raters were asked to keep track of the evidence in the response they considered for scoring each criterion. Discrepancies in rating and in identified evidence were managed through telephone and email interactions with the author prior to further scoring of responses.

The full set of responses was mailed to raters so that each response was scored independently by 2 raters. The author interacted with raters by telephone, email, and in person to provide clarification during the rating process, but not to review ratings.

Analysis

Interrater reliability was analyzed by decision consistency for each criterion. An index of interrater reliability was established by computing the percentage of matching ratings for each criterion.

Evidence of program effects was analyzed using only those criteria for which an acceptable level of interrater reliability was established. For the purposes of this analysis, the level of acceptability was set at 60%. This is a somewhat low level of acceptability for assessment ratings. Because the embedded assessments and scoring were being developed and used for the first time, and because exact matches of ratings were required (due to the three-point scale), this level was deemed to be sufficient for the initial analysis of evidence of program effects.

Rater scores were averaged for each criterion for the analysis of evidence of program effects. Participant and control group scores on the Mathematics Tasks and VideoCase Analyses were compared using repeated measures ANOVA. Participant and control group scores on the Linking to Practice criteria were compared using independent sample t-tests with unequal variances. For the criteria that exhibited significant differences in change over time by group, the statistics are presented below the figure.

Due to the small samples sizes and the pilot nature of the study, adjustments to the alpha level for multiple comparisons were not included. These results, therefore, should be interpreted with caution.

Results

Interrater reliability

The results of the interrater reliability analysis are presented in Table 2. Considering all of the criteria across the six rubrics, 23 out of 45 of the criteria were judged to have a minimally acceptable level of interrater reliability. The nature and wording of these criteria is presented in the analysis of evidence of program effects. In terms of rubrics, the strongest had 4 out of 5 items (Forecasting Student Thinking) and 8 out of 10 items (Exhibiting Disciplinary Content Knowledge) with acceptable interrater reliability. The weakest had 2 out of 10 (Applying Pedagogical Strategies) with acceptable interrater reliability.

Analysis of evidence of program effects: Participant versus control group comparisons

The results of the analysis of evidence of program effects are presented in Figures 1 - 23. The average scores within each group are presented in the graphs. Considering all of the criteria analyzed, several appeared to show a trend favoring the participant group over the control group.

Of these, 7 were judged to be significant through statistical testing (although, again, these results should be interpreted with caution). A trend favoring the control group over the participant group was suggested for a few of the criteria, but none of these were statistically significant. Other criteria appeared to show little notable difference between the participant and control group. Possibly worth noting, however, is that for many of these criteria, an upward trend across both groups was suggested. This trend may be simply a result of practice with the instruments, but could represent some change in learning as a result of doing and reflecting on the instruments. Within this study, it is not possible to determine if either of these explanations may explain these trends.

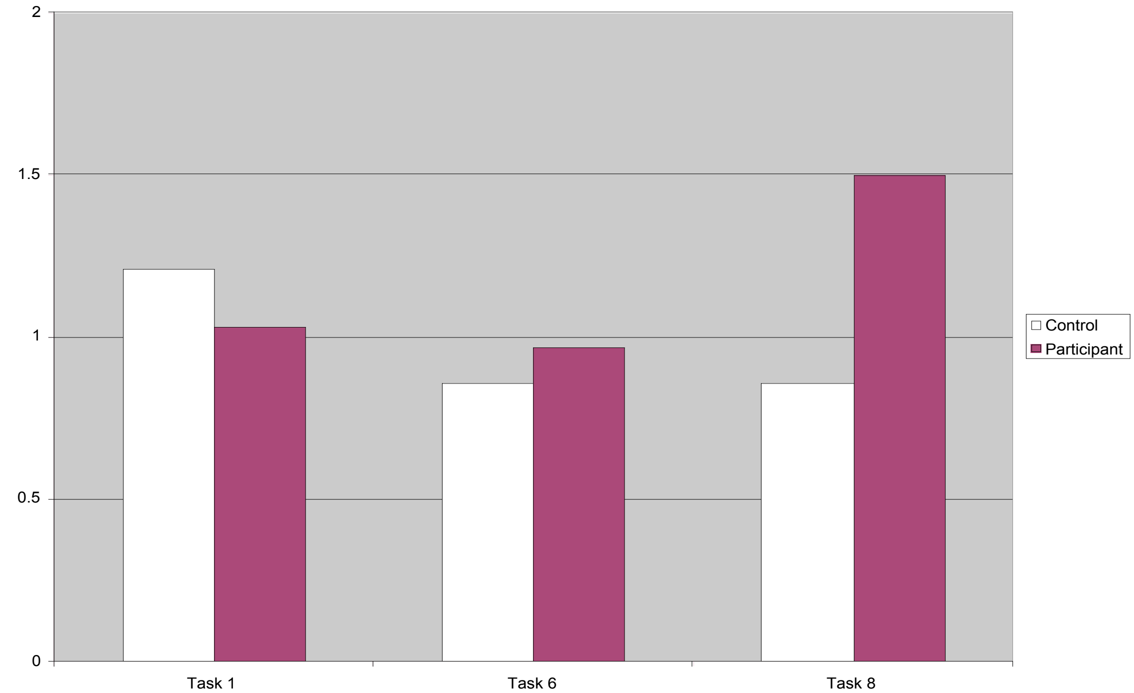
In examining the average scores for some of the criteria, it is apparent that very low average scores were obtained by both participating and control teachers. These results should not be interpreted as indicating a lack of teacher ability to meet these criteria for two reasons. First, some of the criteria, as stated above, relate to potential aspects of the response that were not specifically prompted. These criteria were used in an attempt to assess the propensity of teachers to provide such aspects in their responses. Second, the criteria were not meant to suggest necessary elements of a high quality response to the task. Rather, each criterion was applied in order to examine if the responses contained particular characteristics, regardless of the overall quality of the responses.

Table 2
Interrater Reliability Results for All Rubrics and Criteria

Rubric	Criterion	Interrater reliability
Exhibiting Disciplinary Content Knowledge	MTa1	.56
Exhibiting Disciplinary Content Knowledge	MTa2	.65*
Exhibiting Disciplinary Content Knowledge	MTa3	.86*
Exhibiting Disciplinary Content Knowledge	MTa4	.96*
Exhibiting Disciplinary Content Knowledge	MTa5	.80*
Exhibiting Disciplinary Content Knowledge	MTa6	.60*
Exhibiting Disciplinary Content Knowledge	MTa7	.47
Exhibiting Disciplinary Content Knowledge	MTa8	.65*
Forecasting Student Thinking	MTb1	.72*
Forecasting Student Thinking	MTb2	.54
Forecasting Student Thinking	MTb3	.77*
Forecasting Student Thinking	MTb4	.70*
Forecasting Student Thinking	MTb5	.71*
Selecting Mathematics Tasks	MTc1	.48
Selecting Mathematics Tasks	MTc2	.60*
Selecting Mathematics Tasks	MTc3	.47
Selecting Mathematics Tasks	MTc4	.52
Selecting Mathematics Tasks	MTc5	.70*
Identifying Student/Teacher Thinking	VCa1	.71*
Identifying Student/Teacher Thinking	VCa2	.47
Identifying Student/Teacher Thinking	VCa3	.54
Identifying Student/Teacher Thinking	VCa4	.51
Identifying Student/Teacher Thinking	VCa5	.60*
Identifying Student/Teacher Thinking	VCa6	.36
Identifying Student/Teacher Thinking	VCa7	.75*
Identifying Student/Teacher Thinking	VCa8	.89*
Identifying Student Thinking	La1	.83*
Identifying Student Thinking	La2	.57
Identifying Student Thinking	La3	.54
Identifying Student Thinking	La4	.60*
Identifying Student Thinking	La5	.43
Identifying Student Thinking	La6	.63*
Identifying Student Thinking	La7	.60*
Identifying Student Thinking	La8	.71*
Applying Pedagogical Strategies	Lb1	.49
Applying Pedagogical Strategies	Lb2	.51
Applying Pedagogical Strategies	Lb3	.43
Applying Pedagogical Strategies	Lb4	.49
Applying Pedagogical Strategies	Lb5	.71*
Applying Pedagogical Strategies	Lb6	.57
Applying Pedagogical Strategies	Lb7	.80*
Applying Pedagogical Strategies	Lb8	.54
Applying Pedagogical Strategies	Lb9	.37
Applying Pedagogical Strategies	Lb10	.51

* An index of interrater reliability of .60 or greater was judged to be minimally acceptable.

Figure 1
MTa2*: The response indicates what is the "starting point"/the intercept/static in the pictorial representation of the problem.



*Time by Group Linear Trend: F= 7.57, p<.05

Figure 2
MTa3: The response communicates the relationship between the independent and dependent variables used in the solution.

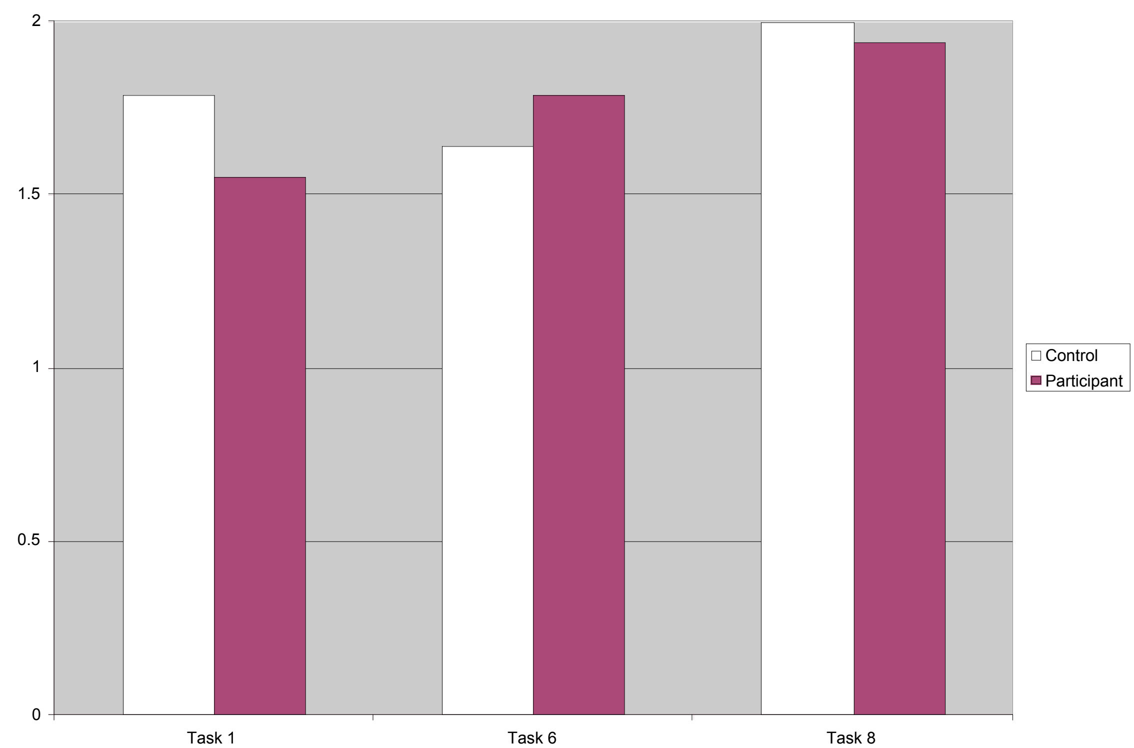


Figure 3
MTa4: The response provides correct solutions [for given numeric values of the independent variable.]

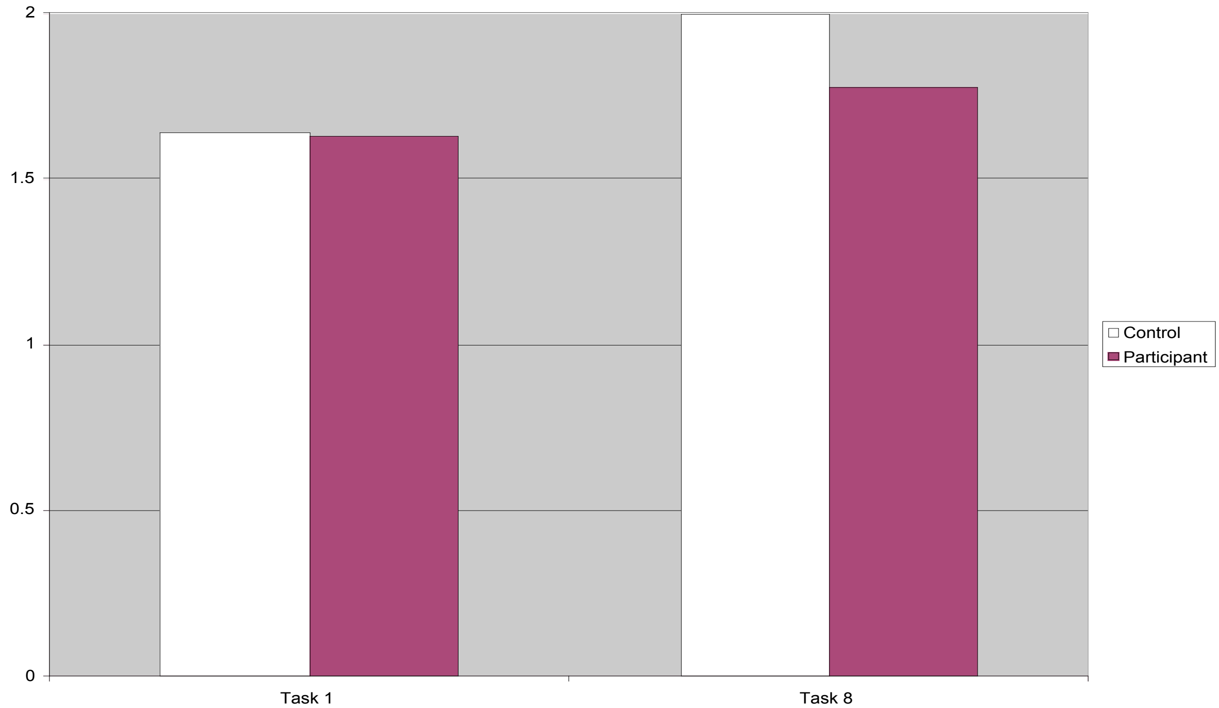


Figure 4
MTa5: The response provides an explicit and correct solution statement for the generalization of the pattern/rule/equation/expression.

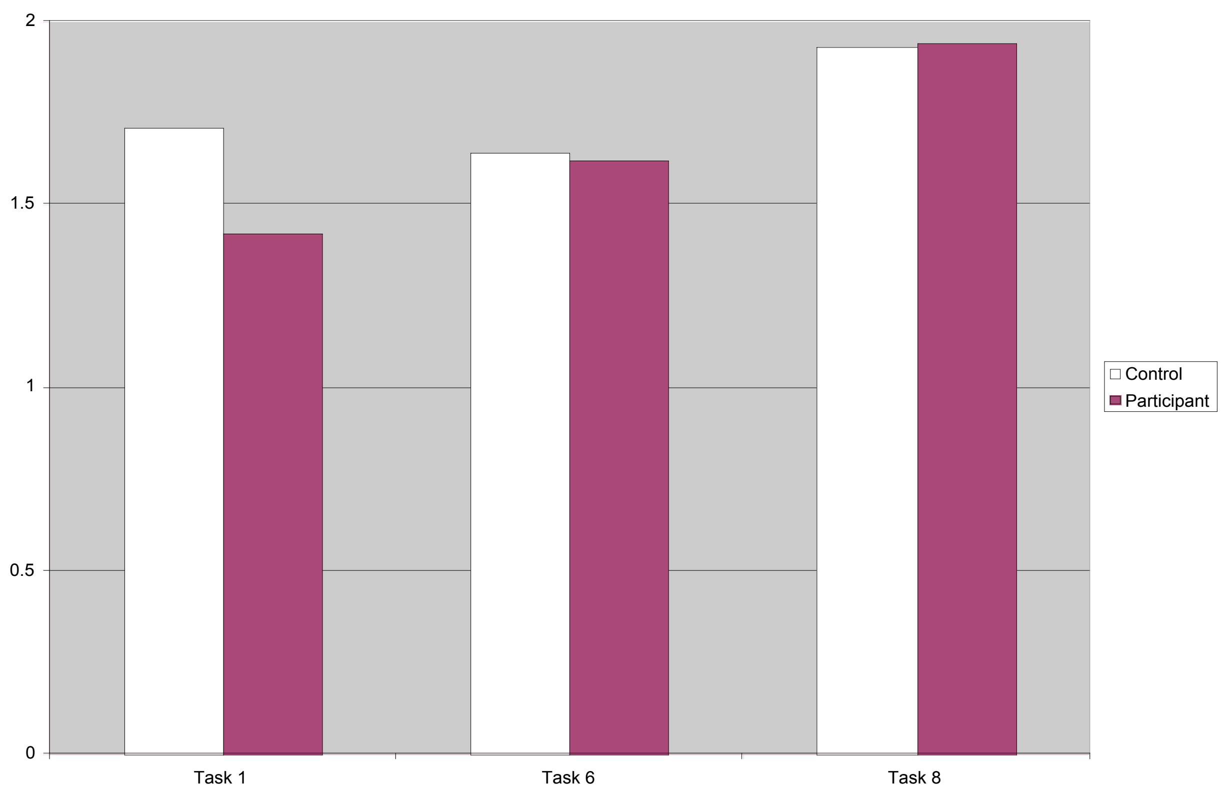
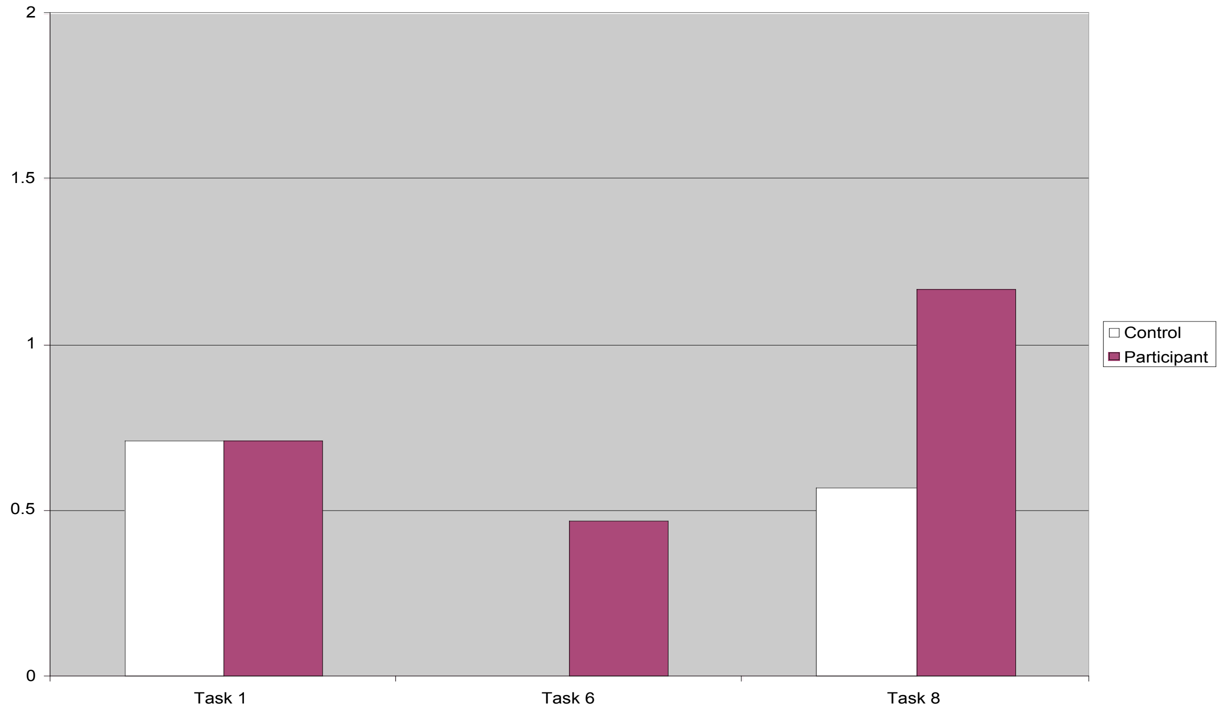


Figure 5
MTa6*: The response provides a description/representation of the arrangement of the [objects represented by the dependent variable].



* Time by Group Linear Trend: F=6.21, p<.05

Figure 6
MTa8: The response indicates a connection between the solution strategy taken to the problem and the "answers" to the specific questions in the solution.

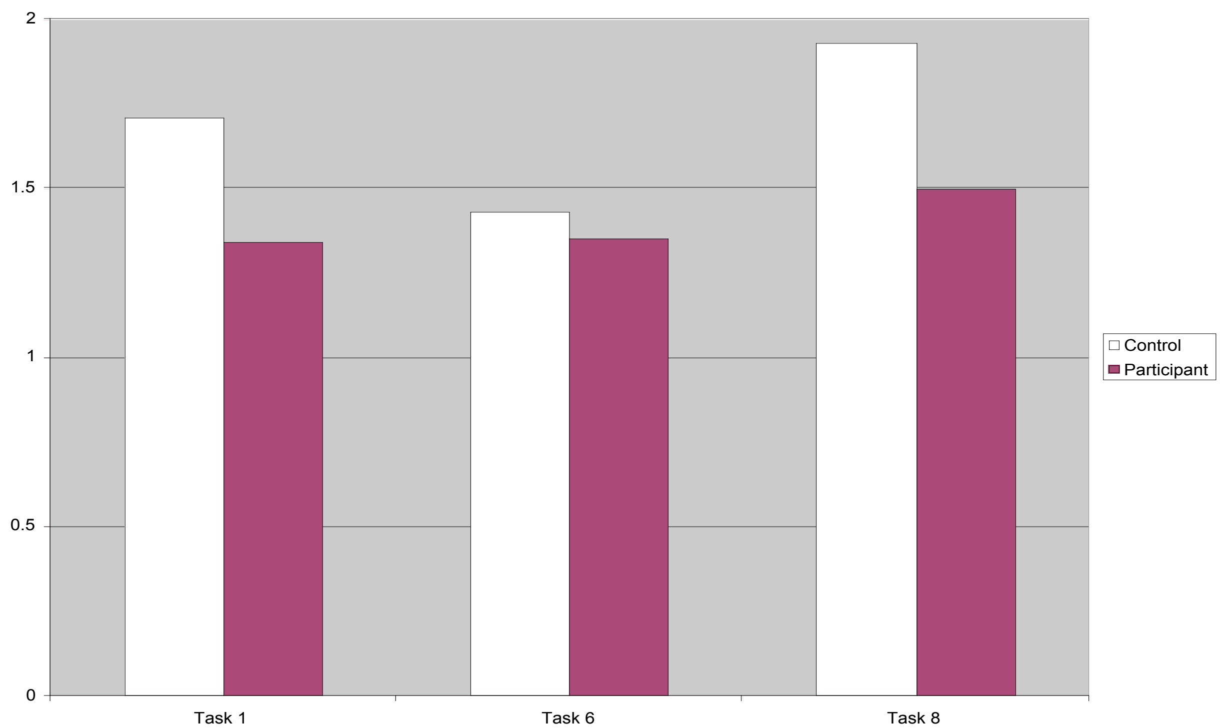


Figure 7
MTb1: The response identifies at least one visualization that is different from the teacher's main approach to the problem and explains the mathematical implications.

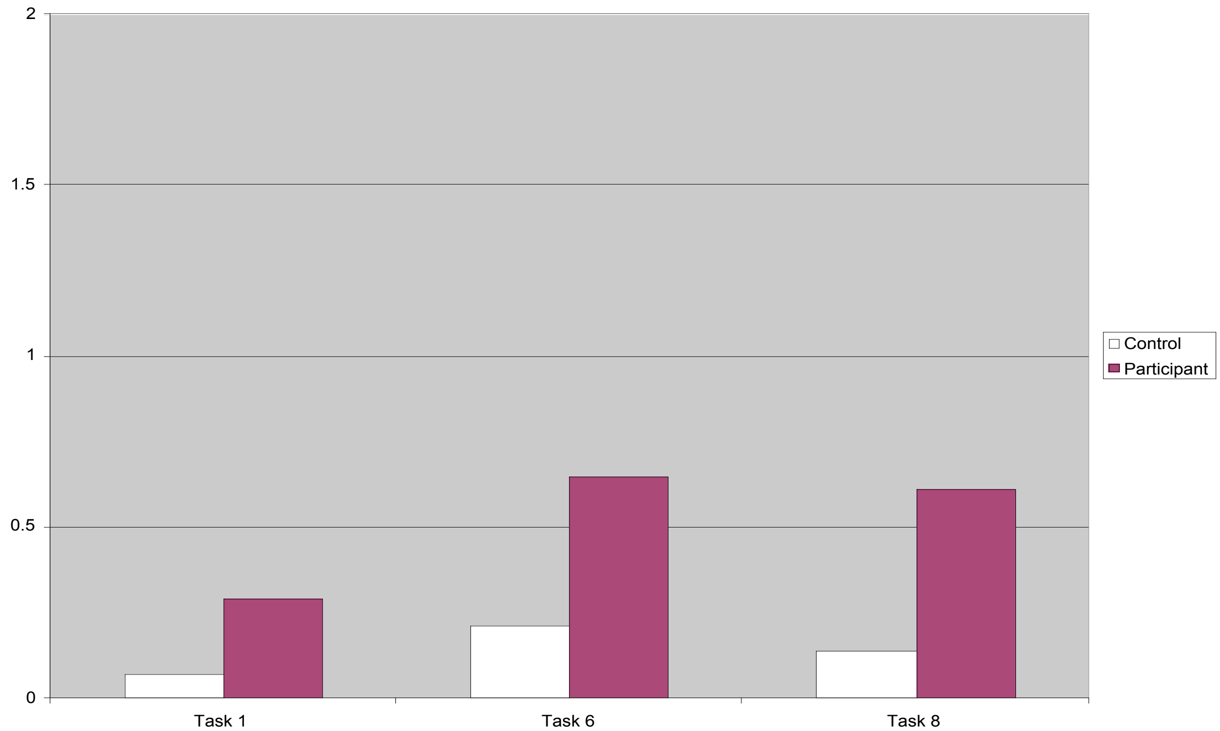


Figure 8
MTb3: The response identifies at least one approach by which students could misrepresent/miscalculate/misinterpret the starting point/intercept of the problem and explains the mathematical implications.

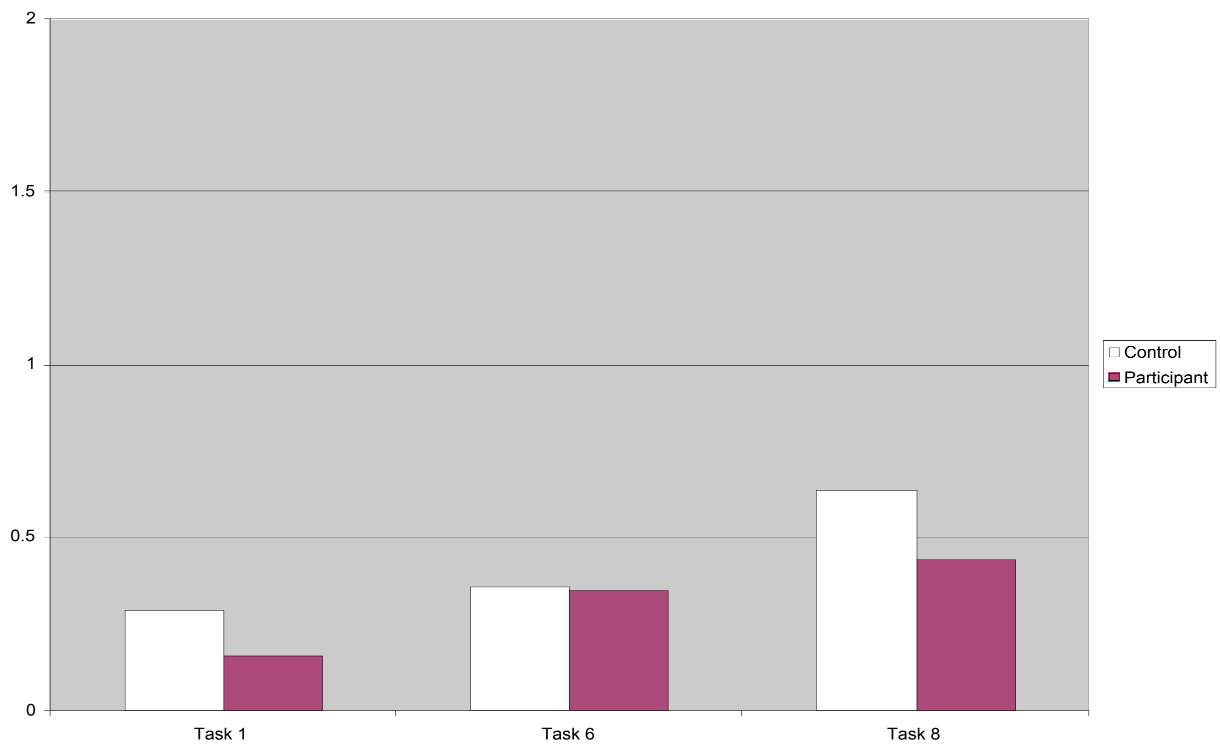


Figure 9
MTb4: The response identifies at least one approach by which students could misrepresent/miscalculate/misinterpret the growth factor/slope of the problem and explains the mathematical implications.

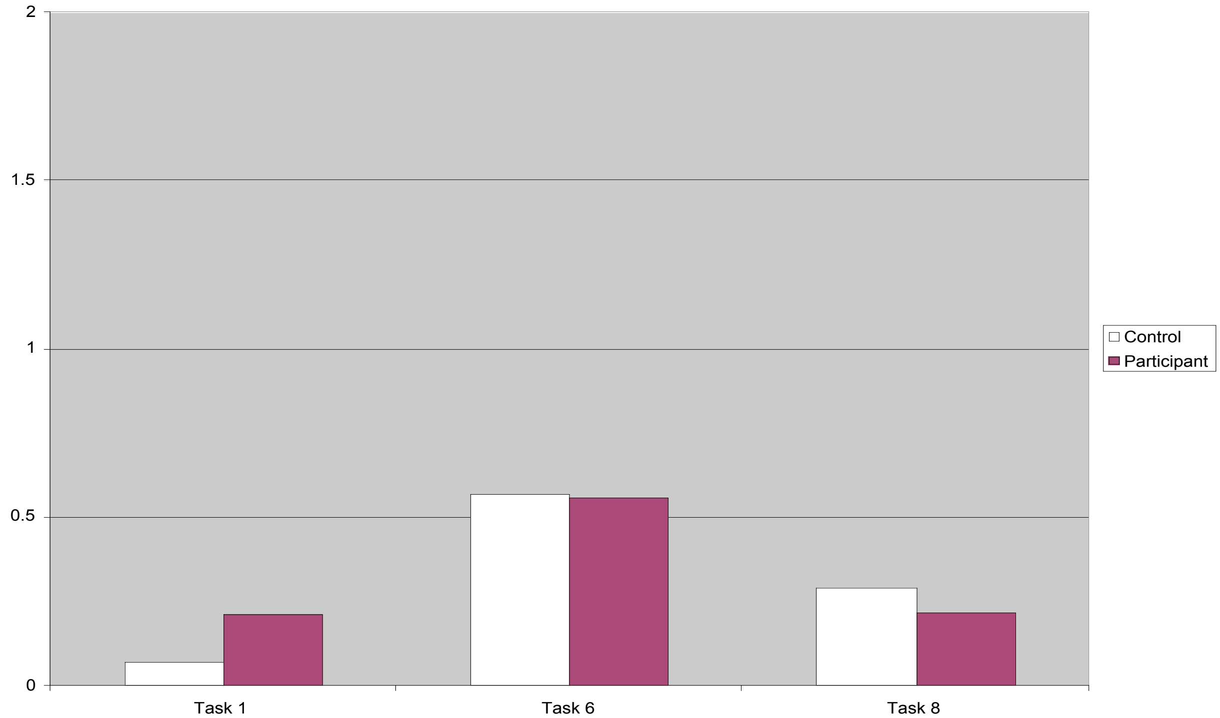


Figure 10
MTb5: The response connects relevant visual features of the pictorial representation of the problem to ways that students might correctly or incorrectly represent the static (starting point/intercept) and/or dynamic (growth factor/slope) aspects of the problem.

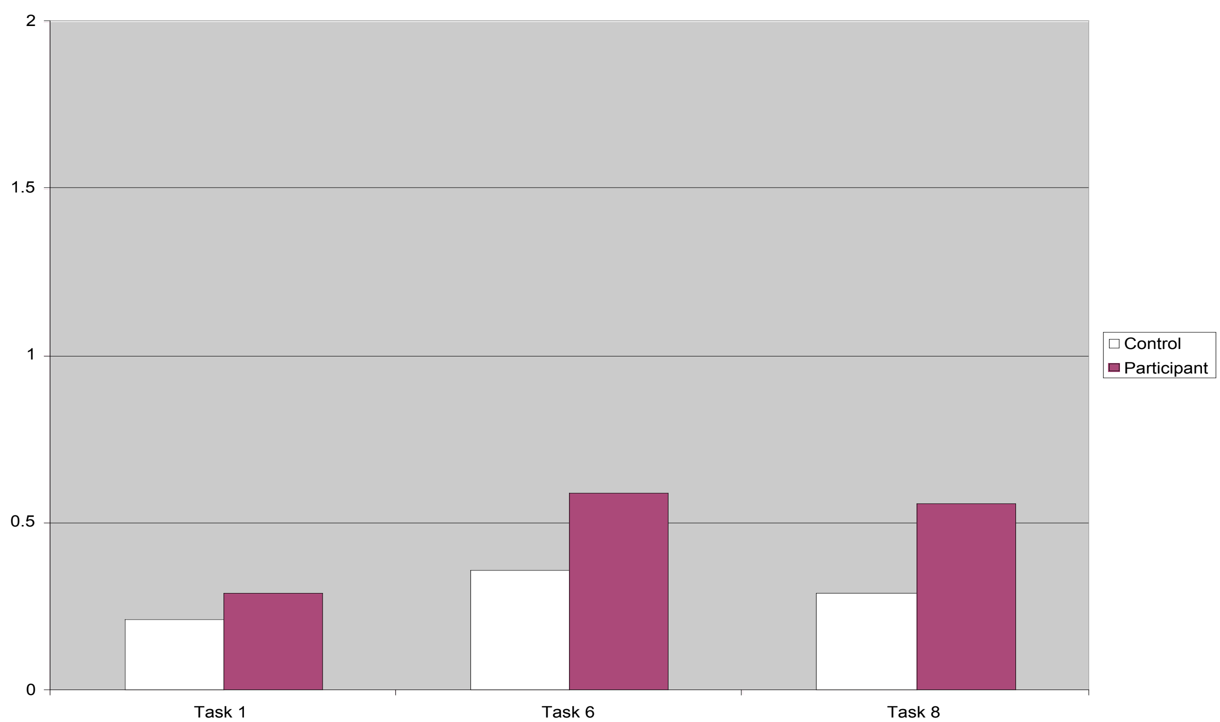


Figure 11
MTc2: The response describes the features of the task that make it useful for teaching/learning the identified concept(s).

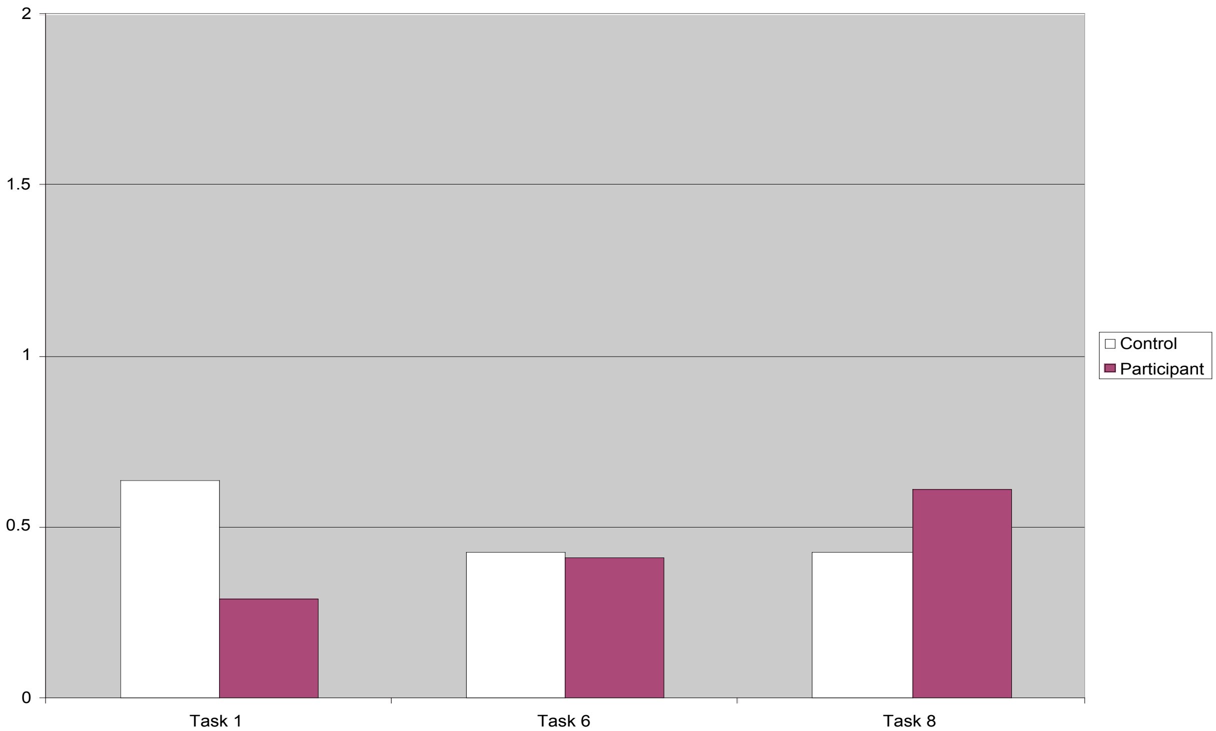


Figure 12
MTc5: The response explicitly identifies a common student learning need that could be addressed using the task.

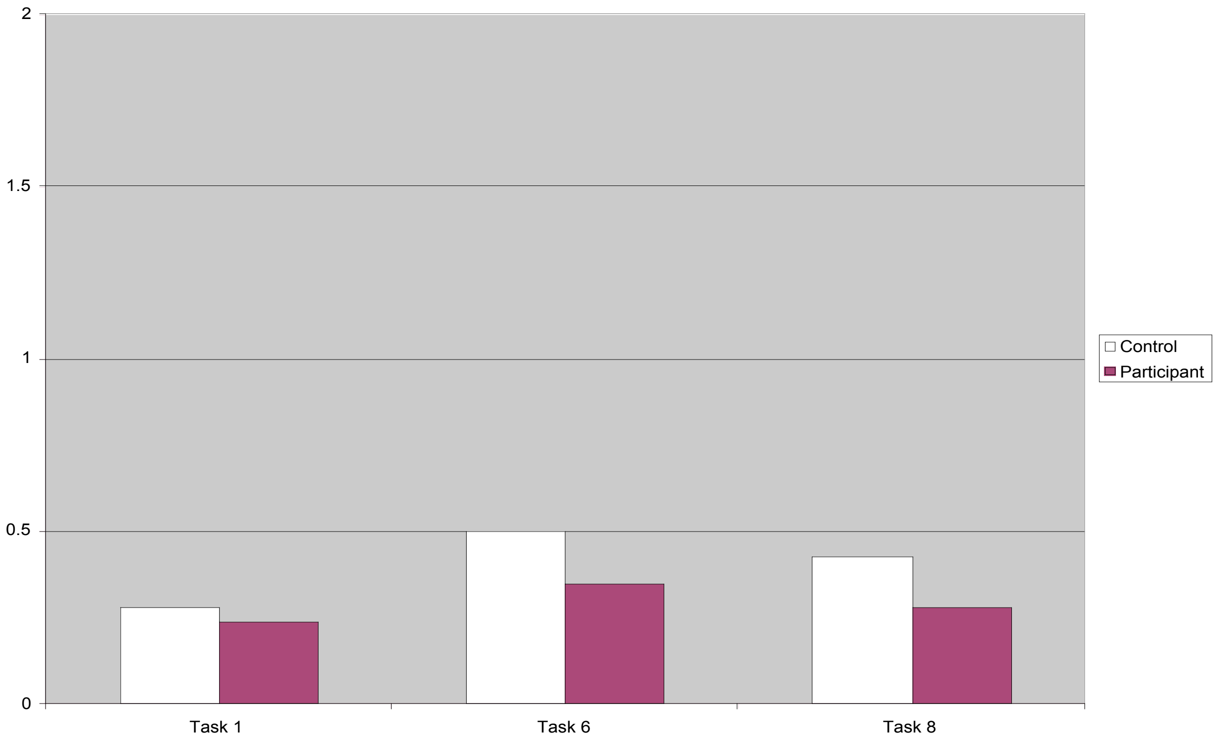


Figure 13
VCa1: The response identifies at least one important mathematics concept/idea that the teacher/students might be thinking about during the identified segment.

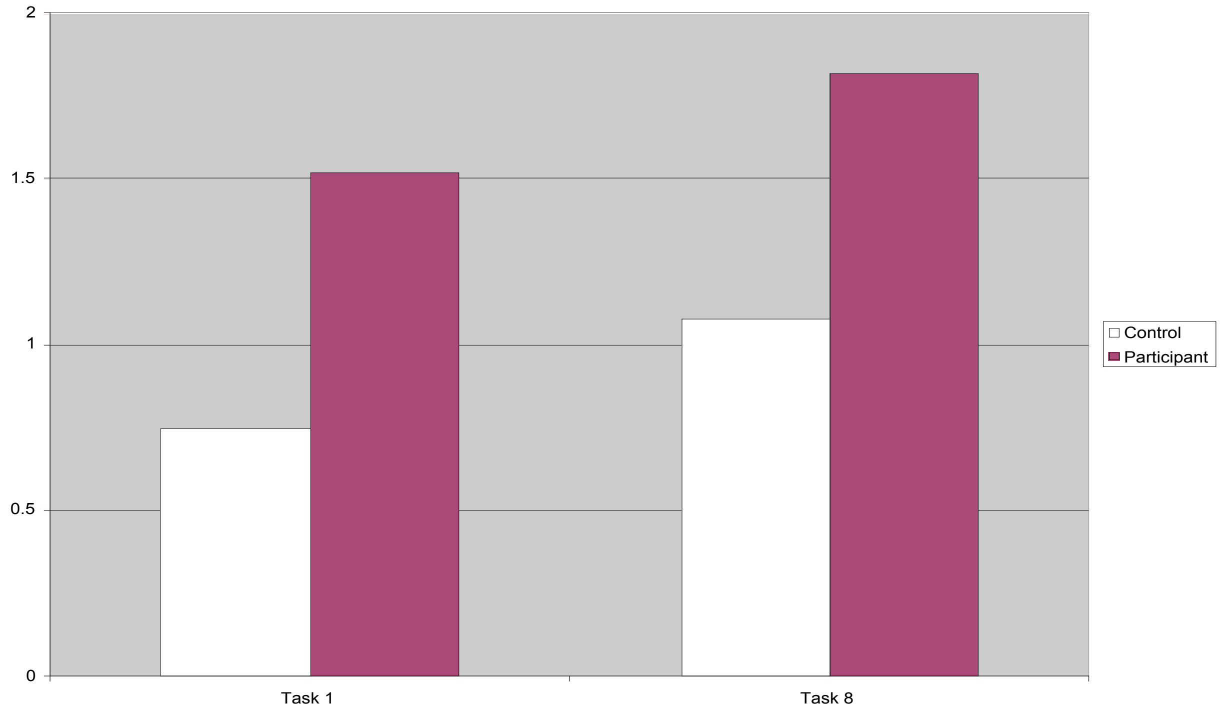
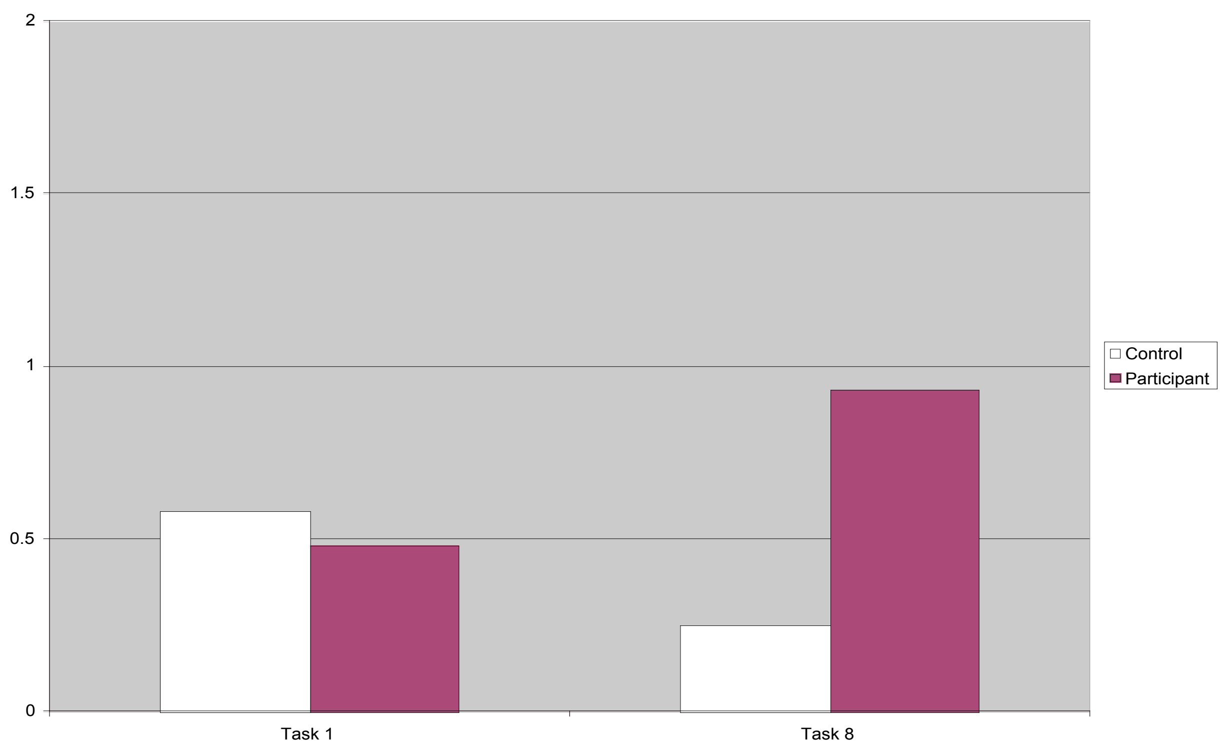


Figure 14
VCa5*: The response provides evidence from the transcript/video to support statements about what the teacher/students might be thinking about the identified concept(s)/idea(s).



*Time by Group Linear Trend: F=5.81, p<.05

Figure 15
VCa7: The response relates the thinking of the teacher/students to a common learning need of students related to the identified mathematical concept(s)/idea(s).

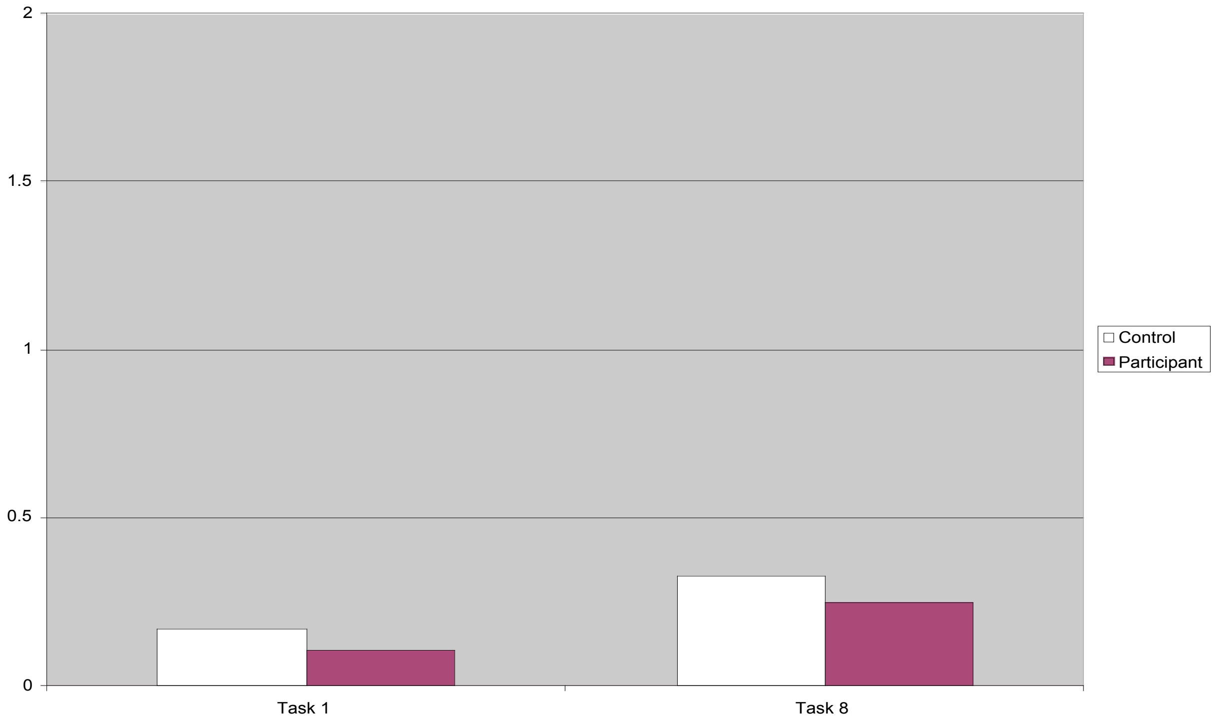
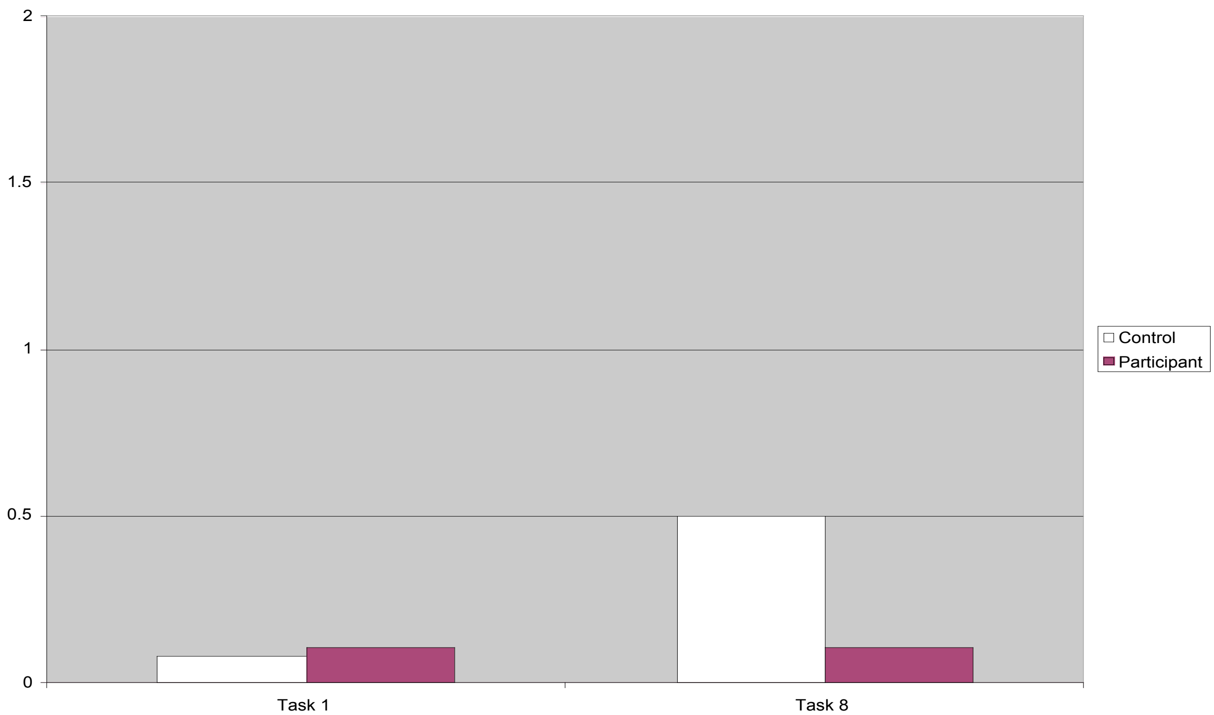


Figure 16
VCa8: The response includes definitive statements about what the teacher should have done in his instruction.



Time by Group Linear Trend: F=5.07, p<.05

Figure 17
LPa1: The response includes an analysis of each students' work individually.

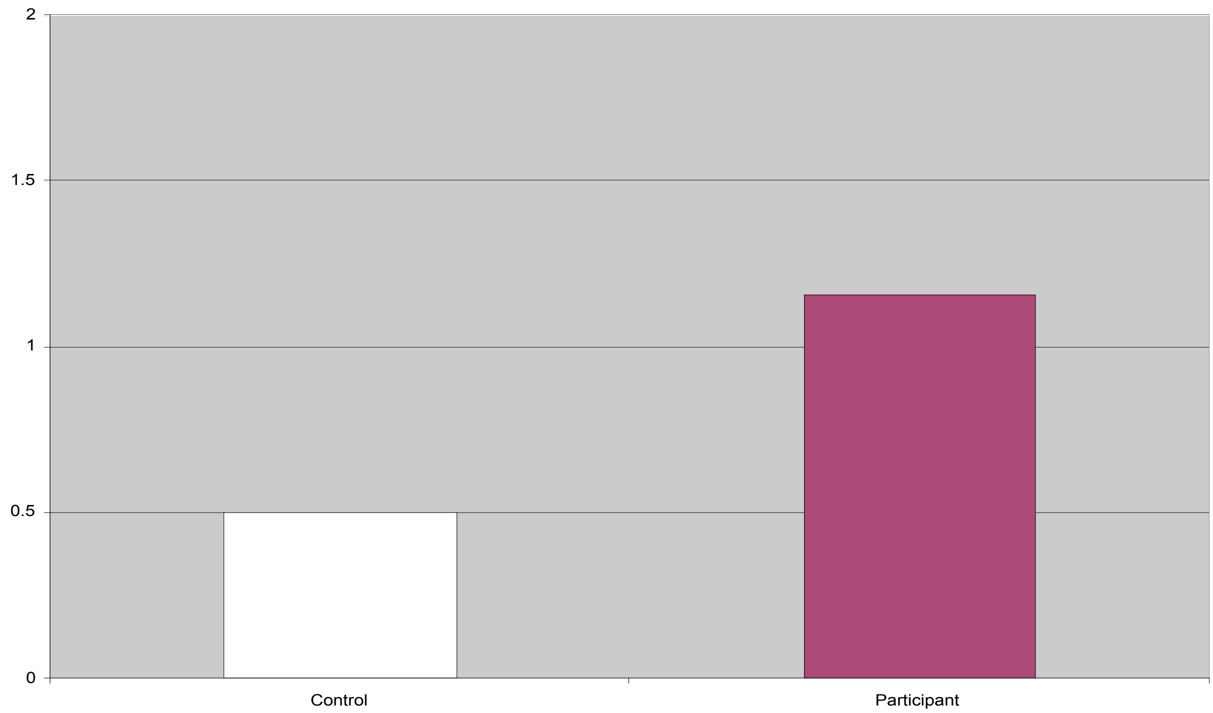


Figure 18
LPa4: The response includes definitive statements about any student's level of understanding of the mathematical concepts of the task.

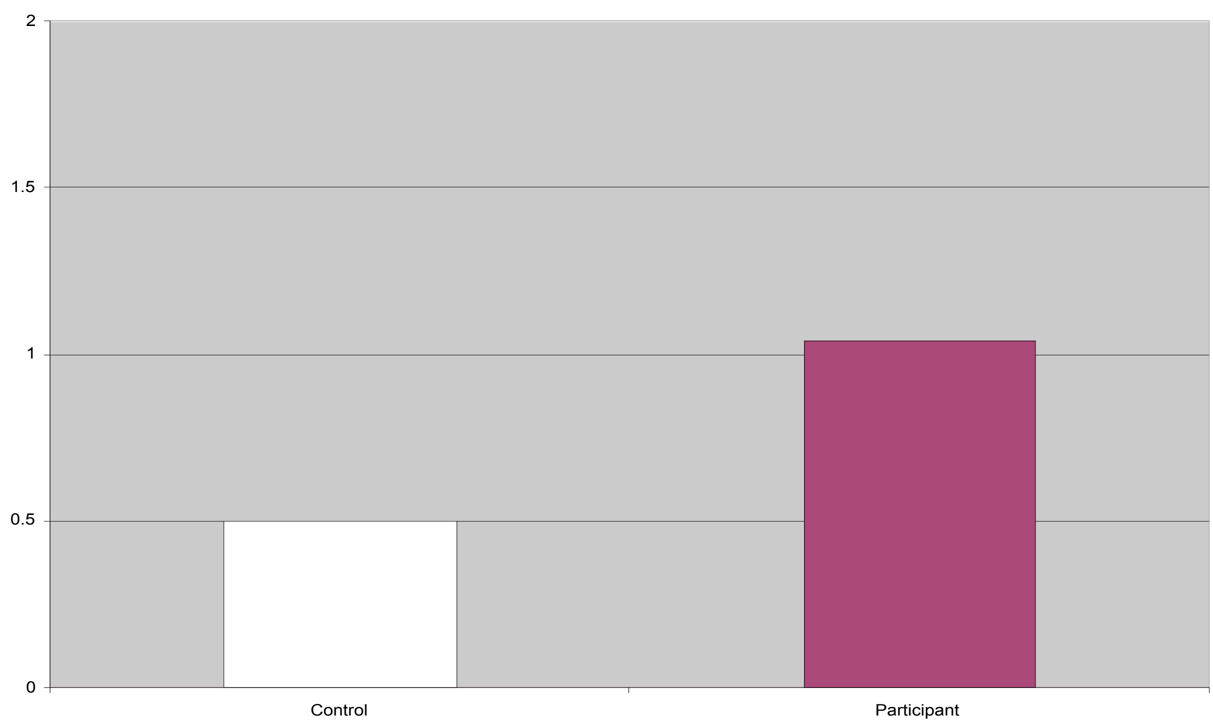
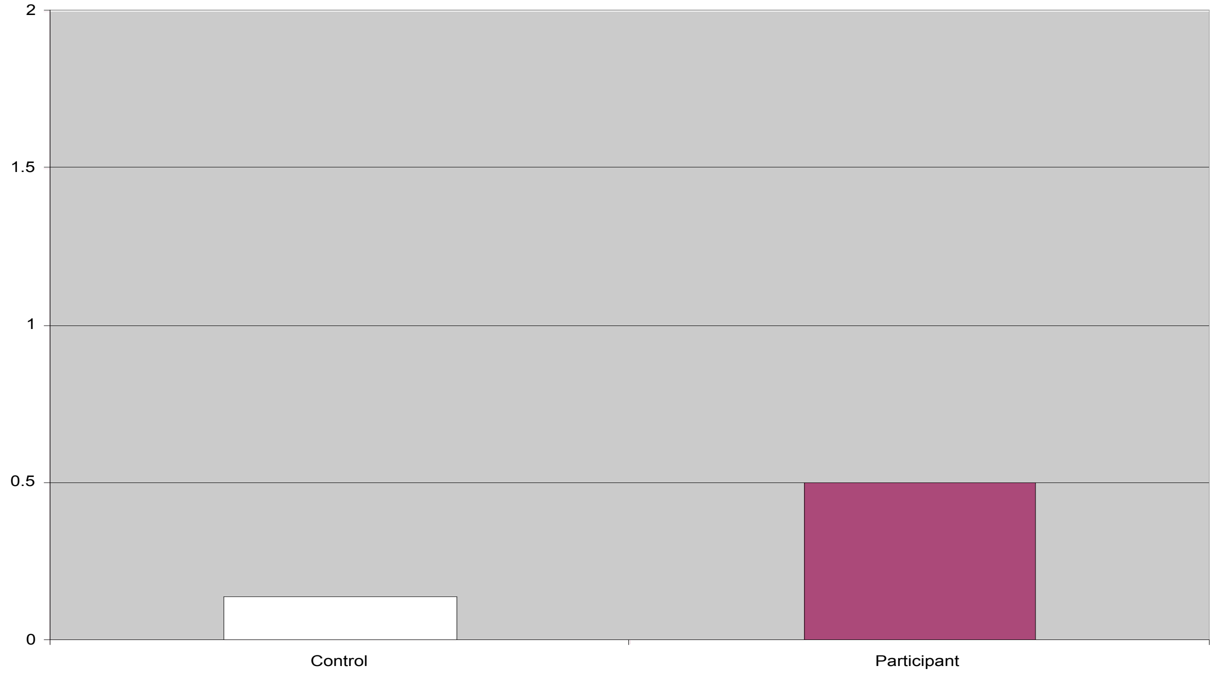
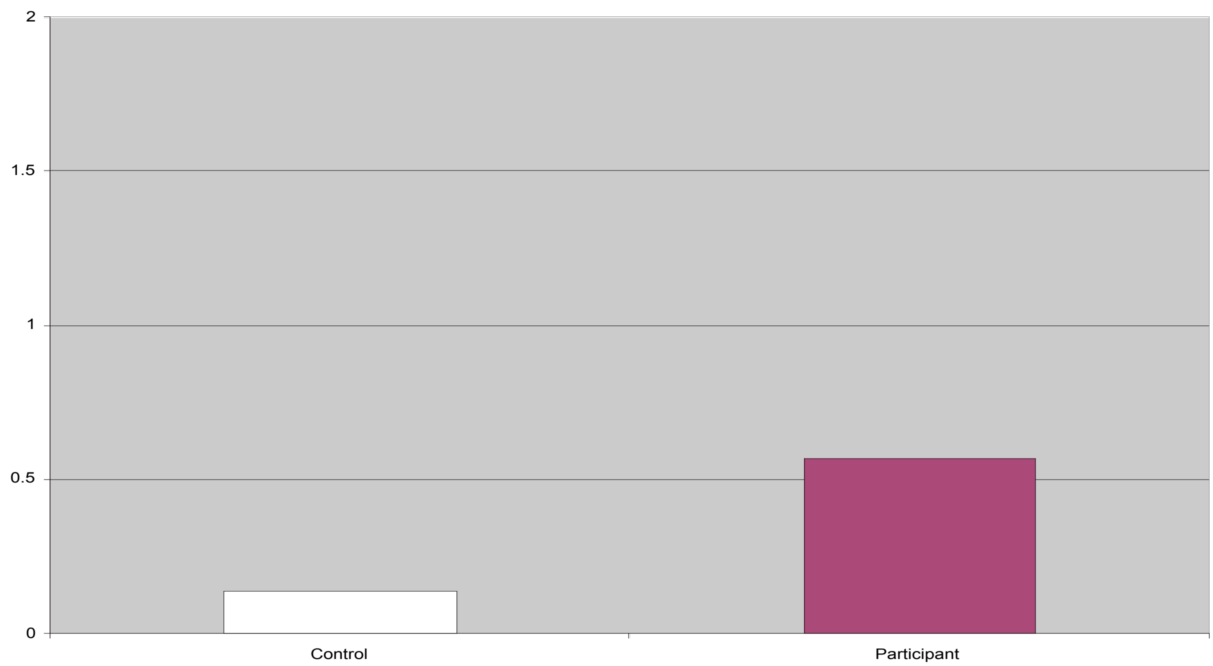


Figure 19
LPa6*: The response provides evidence from the student work to support statements about students' understandings.



*Comparison of means: t= -2.44, p<.05

Figure 20
LPa7*: The response indicates specific aspects of the concept of linear growth that the students have shown evidence of misundertanding.



*Comparison of Means: t=-2.84, p<.01

Figure 21
LPa8: The response provides evidence from the student work to support statements about students' misunderstandings.

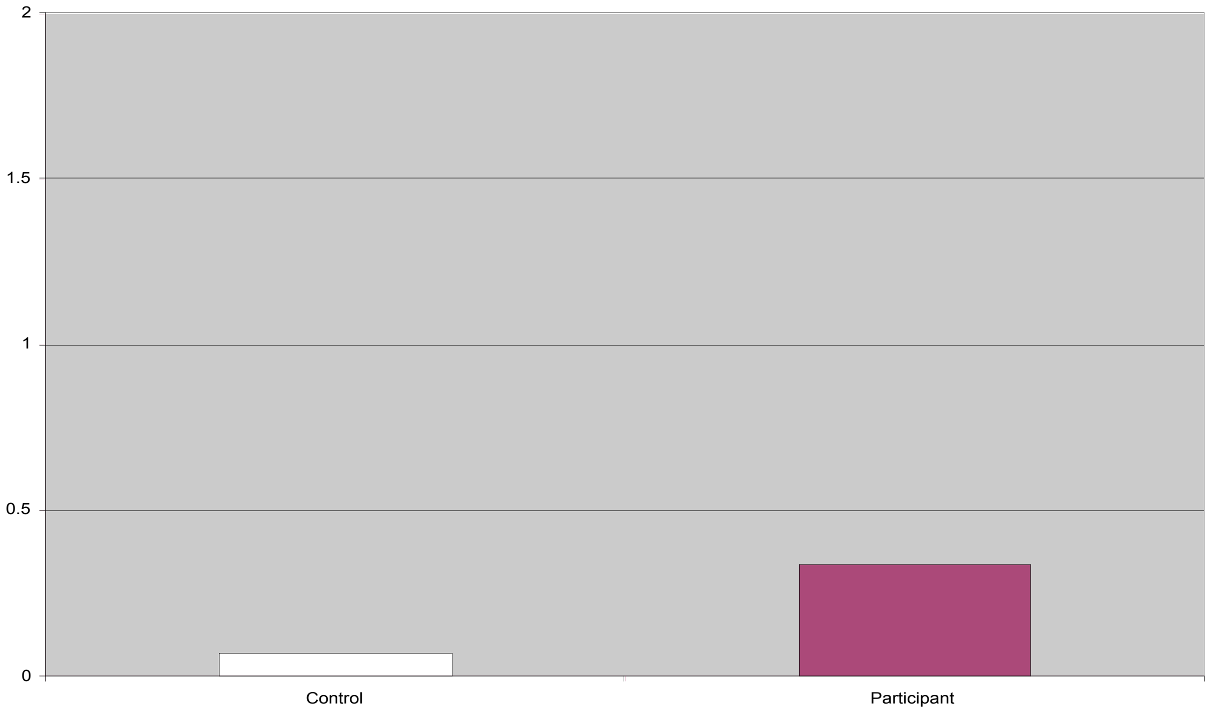
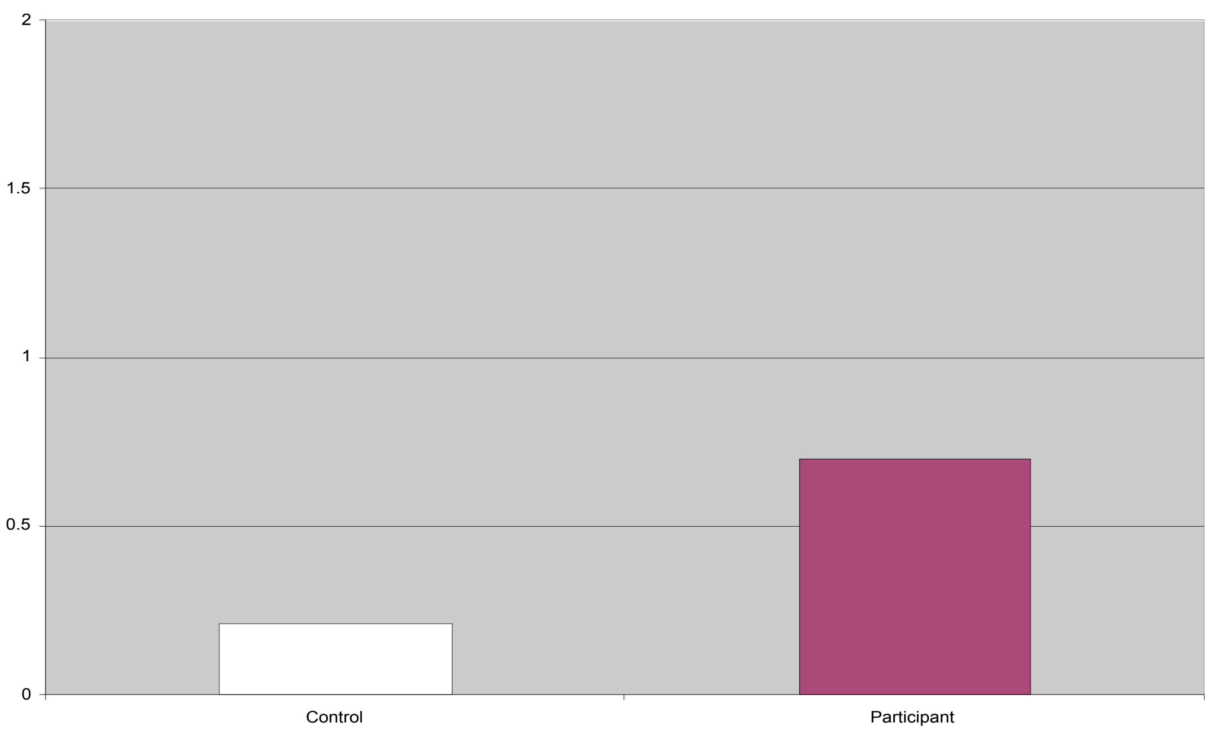
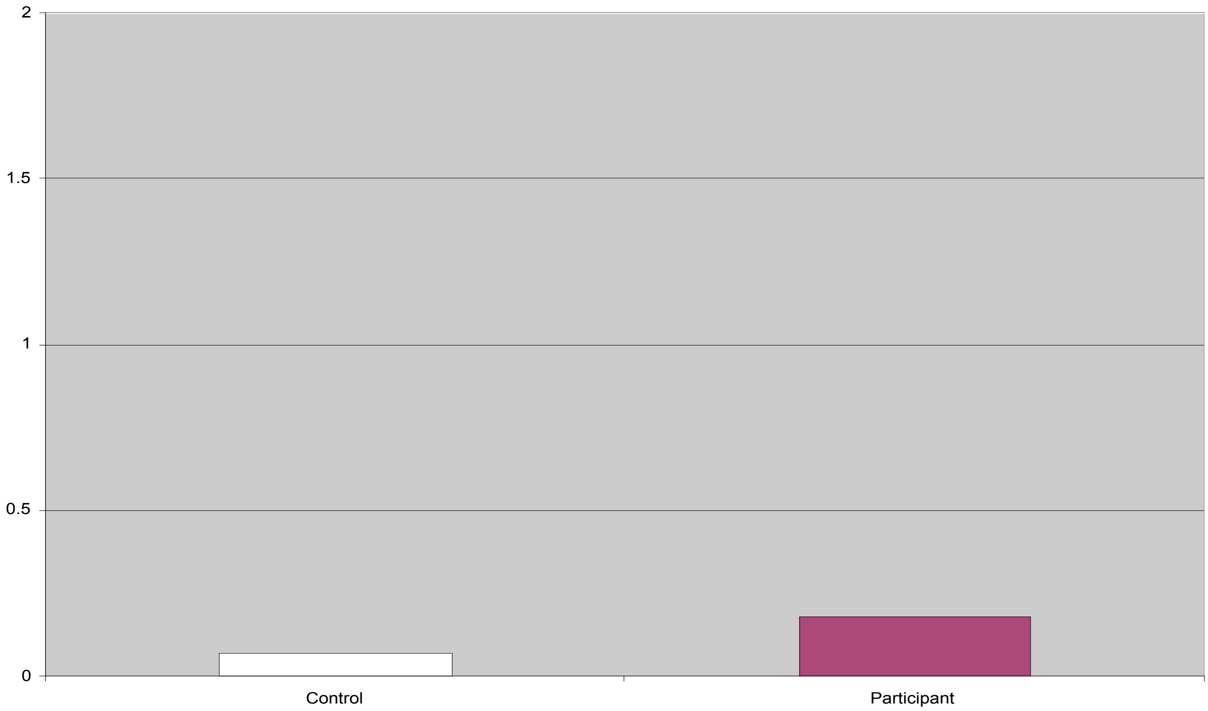


Figure 22
LPb5*: The plan conveys an intention to address students' thinking about the starting point/intercept in the Superstar or a similar problem.



*Comparison of Means: t= -2.88, p<.01

Figure 23
LPb7: The plan conveys an intention to address students' thinking about proportions/proportional reasoning in the Superstar or a similar problem.



Conclusions and Next Steps

The results of the study offer some encouragement to the idea of measuring teacher knowledge in professional development using embedded assessments. Moderate success was achieved in obtaining interrater reliability for ratings of the embedded assessments to measure teacher knowledge, even with a somewhat limited training of raters. The interrater reliability results can inform revision of both the tasks and the rubrics to provide better directions to teachers and to clarify criteria for raters.

In addition to improving the measurement properties of the individual criteria through refinement of the instruments, rubrics, and scoring process, an important next step will be the creation of scales from the instruments. Although the rubrics were developed to include sets of criteria related to one of the learning goals of the VCMPD program, and some of the criteria identify related abilities/propensities, the pilot phase did not allow for scale development due to the foregoing need to establish interrater reliability and a very small sample size for factor analytic techniques to be applied. In the interest of examining the instruments at the criterion level, the number of contrasts presented in this study is admittedly large. The development of conceptually and statistically coherent scales that combine related criteria will improve the measurement and analysis as work in this area progresses.

The moderate, but generally positive, results of the analysis of evidence of program effects offers some support to the idea that the embedded assessments are measuring teacher knowledge of the learning goals of the VCMPD program, and that teachers are indeed increasing their ability and/or propensity in targeted areas.

In terms of exhibiting mathematical knowledge pertinent to the teaching of mathematics, VCMPD participants were statistically more likely than control group teachers to increase in their ability/propensity to connect their work on the Mathematics Tasks back to the pictorial representations in which the task originated. This finding was supported both through identifying the intercept in the pictorial representation and through describing or representing the arrangement of the pictorial representation in their solution. Aiding students in translating from pictures to more abstract representations, such as equations, likely depends on teachers' ability and propensity to make such connections.

The increase in attention to the intercept as represented in the pictorial representation may reflect another effect of the program, as participating teachers were more likely than control group teachers to attend to this concept in their plans for instruction on the Linking to Practice instrument.

On the VideoCase Analysis instrument, participating teachers increased more than control group teachers in their ability/propensity to provide evidence from the video or transcript to support their statements about what the teacher or students might be thinking about the mathematics of the videocase. Similarly, on the Linking to Practice instrument, the VCMPD participants were statistically more likely than control group teachers to provide evidence from the student work to support their statements about students' understandings of the targeted concepts of the task, and to indicate specific aspects of the concept of linear growth that the students showed evidence of misunderstanding. Citing evidence to support claims and statements, and being specific about mathematical aspects of artifacts in written work and case discussions are encouraged in the VCMPD program. These practices are viewed as good habits of reflective and investigative professional development for teachers. It is also an objective of the program to develop teachers' ability/propensity to think very specifically about what students say and do when making assessments of their mathematical progress or understanding.

Also on the VideoCase Analysis instrument, control group teachers were somewhat more likely to increase in their propensity to indicate definitively what the teacher depicted in the video should have done in his instruction. The VCMPD program stresses that the videocases are objects for teacher inquiry and reflection, and represent only a very small sample of interaction between a teacher and his/her students. The program strongly cautions against making definitive statements about the practice of the videotaped teachers based on so small a sample of instruction and almost no contextual information.

The reader should again be cautioned about interpreting the results of this pilot due to concerns about interrater reliability across all rubrics and criteria, and low sample size, especially of the control group. However, the suggestion of the findings is that the VCMPD program does have the intended effects on some of the targeted teacher abilities/propensities beyond what can be accounted for by practice on the tasks or a "volunteer effect" of teachers who are likely to show interest in the VCMPD program.

Furthermore, the findings offer modest support to the use of embedded assessments to measure the targeted learning goals within a professional development context. In addition to revising and improving the instruments and the scoring rubrics, training, and procedures, an important next step is to trace the measured teacher knowledge gains further along the logic model presented earlier. It is a very important step in our understanding of professional development to learn and document what teachers gain in their abilities and propensities within professional development.

The next critical step in examining the theoretical underpinnings of practice-based professional development is to determine how and to what extent teacher knowledge gains in professional development translate into classroom practice.

Finally, this study has largely limited the examination of embedded assessments to summative evaluation uses. A number of formative uses have also emerged or have been suggested by this work. First, embedded assessments can be a useful tool in teacher education materials development. Collecting the kind of information the embedded assessments provide during piloting and field testing of teacher education materials offers developers a window into how teachers experience the materials and what they might or might not be learning. Second, the embedded assessments and rubrics might offer a means to operationalize teacher learning goals for professional development providers who will use the program with teachers. The kind of teacher learning the program targets could be represented and investigated through sample responses on the embedded assessments. Third, embedded assessments could offer a means for professional development providers to diagnose particular teacher learning needs in professional development and to identify areas of strength from which to build new understandings. Fourth, the embedded assessments, and the rubrics for scoring them, might offer a way for professional development providers to investigate teacher work as a part of their own professional growth. That is, the embedded assessments can provide a means for examining artifacts from professional development just as practice-based professional development often focuses on artifacts from the classroom. Fifth, embedded assessments might be used by teachers in professional development to reflect on their own learning, for example by examining their responses to the embedded assessments as a record of their thinking over time. It may, in fact, turn out that embedded assessments to measure teacher knowledge in professional development offer much more for formative than summative purposes.

References

Ball, D. L. and. Hyman, B., (2000). Interweaving content and pedagogy in teaching and learning to teach: knowing and using mathematics. In J. Boaler. (Ed.), *Multiple Perspectives on Mathematics Teaching and Learning* (pp. 83-104). Westport, CT: Ablex Publishing**.**

Carpenter, T.P., Fennema, E., Peterson, P.I., & Carey, D.A. (1988). Teachers pedagogical content knowledge of students' problem solving in elementary arithmetic*. Journal for Research in Mathematics Education, 19*, 385-401.

Fenemma, E., & Franke, M.L. (1992). Teachers' knowledge and its impact. In D. Grouws (Ed.) *Handbook of Research on Mathematics Teaching and Learning* (pp. 147-165).. New York: Macmillon.

Franke, M.L., Carpenter, T.P., Levi, L., & Fennema, E. (2000) Capturing teachers' generative growth: A follow-up study of professional development in mathematics*. American Educational Research Journal, 38*, 653-689.

Kennedy, M. (1998). *Form and substance in inservice teacher education. Research monograph no. 13*. Madison, WI: National Institute for Science Education.

Ma, L. (1999). *Knowing and Teaching Elementary Mathematics*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Putnam, R. T., &; Borko, H. (2000) What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4-15.

Wilson, S.M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: A review of research on contemporary professional development*. Review of Research in Education, 24*, 173-209