Factors Influencing Middle Grades Students' Algebra Learning: Multiple Research Perspectives

Linda Dager Wilson, Victor L. Willson, Jon Rahn Manon, Ross Larsen, Yan Li, and Jo Ellen Roseman

Paper presented at the
Annual Meeting of the American Educational Research Association (AERA), April 14, 2009, in San Diego, California

Introduction

This paper presents findings from research on the teaching and learning of the algebraic concept of change, applying multiple research perspectives to the same data base. The data set is longitudinal, following teachers in grades 7 and 8 for three years, and unique, in that it links curriculum materials (four commonly used textbooks) to specific teaching strategies to student outcomes on the same learning goal. Quantitative findings, using multilevel structural equation modeling, are examined more closely via case studies. We describe the study, the data collected and measures used, results from the quantitative analysis, and the subsequent case studies.

Background

An Interagency Education Research Initiative (IERI) project jointly funded by the Department of Education, National Institutes of Health, and the National Science Foundation, REC-0129398, was awarded to the American Association for the Advancement of Science, in collaboration with the University of Delaware and Texas A&M University to study the impact of professional development on instruction and student achievement for specific learning goals in middle school mathematics. The IERI study was based on Project 2061's earlier studies of the quality of textbooks' support for middle school mathematics teaching and learning (American Association for the Advancement of Science [AAAS], 2000).

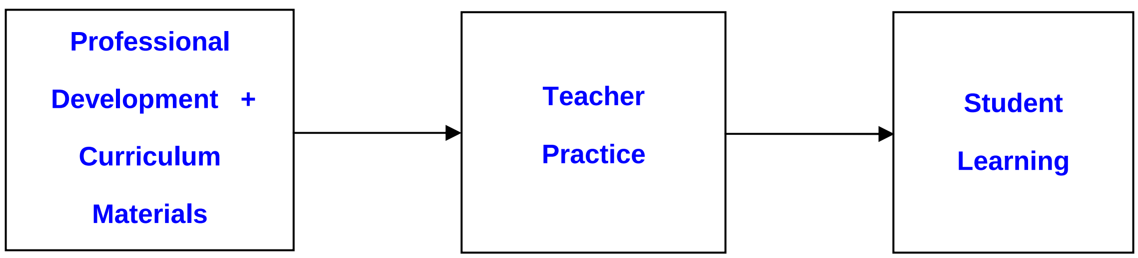
Research Question

The IERI study was designed, in part, to identify factors that contribute to middle school students' learning about the algebraic concept of change. The factors examined were instruction, curriculum, and professional development. These factors, as well as the measures of student learning, were each considered relative to the following learning goal for grades 7 and 8:

> ***Symbolic equations can be used to summarize how the quantity of something changes over time or in response to other changes.***

This particular learning goal was chosen for its centrality in the algebra curriculum for 7th and 8th grades (Kieran, n.d.; Kaput, 1999). The study was designed so that this learning goal would serve as a common filter for each potential factor, enabling a close study of multiple teachers using multiple materials but teaching a common concept. Holding the mathematical content constant provided an opportunity to increase the sensitivity of the measures used to examine the potential factors that might influence learning.

One of the hypothesized relationships central to the study is that certain features of instruction would make a significant difference in student learning. We also examined the extent to which different curricular lessons, each offering a different pedagogical approach and varying degrees of support for instruction, predict student learning. In addition, the teachers participated in professional development activities during the three years of the study, and we examined the extent to which this activity predicted teaching practice and student learning. The logic model below shows each of the factors considered in the study:



Professional Development. Teachers participated in professional development jointly planned and carried out by researchers at AAAS, the University of Delaware, and Texas A&M University over three summers and during each of three academic school years. The focus of the professional development was the algebraic concept of change between two variables, with the intended goal of understanding how it can be represented mathematically and how students might show evidence of learning about it. During the first year teachers considered the meaning of the learning goal and identified the specific lessons in their own curriculum that targeted that idea, but beyond that the activities were not directly related to any of the four curricula used by the participating teachers. The second year's activities dealt with the use of multiple mathematical representations of algebraic change, and in the third year the focus shifted to finding evidence of student learning in videotapes of project classrooms or in students' written work.

Curriculum. The four curriculum materials used in the study were: *Connected Mathematics* (Lappan et al., 1998; Lappan et al., 2000), *Mathematics in Context* (Romberg et al., 1998), *Middle Grades Math Thematics* (Billstein et al., 1999), and *Mathematics Applications and Connections* (Collins et al., 1998). Each of these materials included a set of lessons (3-5) that targeted the learning goal on algebraic change. The teachers were not directed on the use of their given materials, and fidelity to those materials varied. Some teachers, particularly those using either

Math Thematics or *Mathematics Applications and Connections*, supplemented the lessons freely with handouts or other materials, while the teachers using *Connected Mathematics* and *Mathematics in Context* tended towards a stricter adherence to the lessons as written.

Instruction. Participating teachers taught the specified lessons from their own curriculum, typically lasting three to five days. Our hypotheses about the aspects of instruction that might make the greatest difference in student learning were based on research literature, and on the set of 24 criteria used in the AAAS textbook analysis (AAAS, 2000). From that set we convened an expert panel to choose five of the textbook criteria that they thought were most central to learning and adapted them to examine teaching. These five were:

1. Representing Ideas Effectively (IV.C): Does the teaching use accurate and accessible representations of the learning goal?

2. Encouraging Student Explanations (V.A): Does the teaching encourage each student to express, clarify, justify, interpret, and/or represent his/her ideas about the learning goal?

3. Asking Guiding Questions (V.B): Does the teaching use questions and/or tasks to guide student interpretation and reasoning about experiences with real world examples, representations, and/or readings related to the learning goal?

4. Finding Out Students' Ideas (II.C): Does the teaching use questions and/or tasks to find out what students think about familiar situations related to the learning goals <u>before</u> these goals are introduced?

5. End of Lesson Assessment (VI.B): Does the teaching engage students in assessment questions and/or tasks that require students to show, use, apply, explain, or otherwise demonstrate their understanding of the knowledge and/or skills specified in the learning goal?

One of the aims of the study was to see which, if any, of these instructional strategies might predict student learning.

Methods

Subjects

Students. A total of 2,484 grade 7 and 8 students from whom permission forms were returned participated in taking the pre-test, the post-test, or both over the three year period: 1,025 in year 1, 861 in year 2, and 598 in year 3. (Note that while the teachers stay constant, the students do not.) A total of 1,696 took both pre-test and post-test. Failure to take both was primarily due to absence on the testing date. There were 1,995 grade 7 and 529 grade 8 students in the sample. There were no grade 8 students in the first year of the study because the study design called for only grade 7 teachers in the first year, with grade 8 added in the second year. Boys and girls were approximately equally represented. Ethnicity was not coded, although based on school district reported results, the Texas students were tri-ethnically mixed in about equal numbers, while in Delaware there were mostly white or African-American students.

Teachers. A total of 28 grade 7 teachers and 13 grade 8 teachers participated in one or more years of the study. Student data across all three years was available from 12 of the participating teachers, two years of data from 15 teachers, and one year from 14 teachers. Average teaching experience was about 9.5 years, ranging from 0 to over 30 years. All but five of the teachers were female.

Measures

Variable	Measure(s)
Student achievement	Post-test (adjusted for Pre-test) by item, item type, and whole test
Teaching quality	Proportion of class time focused on learning goal and quality of use of research-based instructional strategies targeting it.
Curriculum quality	High or not high AAAS rating
Professional development	Number of days/year over 3 years and number of years

Student achievement. We developed student assessments for the learning goal, beginning with a concept map of mathematical ideas encompassed by the learning goal (Appendix A). We identified three main constructs that contribute to understanding of the learning goal: Change, Variables, and Equality and Equations. From the map, an achievement test was developed that consisted of 8 multiple choice items and 8 constructed response items. One item, termed the “superitem” (Collis, Romberg, & Jurdak, 1986) was a four-part item intended to evaluate students’ higher level thinking about the learning goal, while optimizing access for students. Three of the constructed response items were scored using rubrics that gave 0-2 points, as were three of the four parts of the superitem. Thus, the total score that could be achieved on the test was 25 points. Parallel forms of the test were constructed for pre-test and post-test administration. Reliability of the post-test exceeded .8 for both forms. The pre-test exhibited comparable reliability, but this was not a criterion, since most students had not previously been presented with the lesson concepts and reliability for the pre-test was primarily an indicator of previous knowledge or mathematical ability rather than instructional variation.

The algebra pre-test was given sometime prior to the lesson instruction, but the time of administration varied across teachers. Similarly, the post-test was administered from weeks to months after the lesson at the direction of the teacher. No attempt was made to adjust for the differential times, and it was considered a random variable that would not systematically affect any specific factors in the data analysis.

Factor analysis of the post-test was conducted after year 1 and provided only modest support for the original conceptual structure. The structure is not the focus of this paper, and detailed analysis is in progress to demonstrate changes over time. Since the total score on the test was used as the criterion measure to evaluate instructional effects, the exact composition reflects the actual knowledge structure developed by students. There was evidence that this structure was significantly different across curricula. It is asserted here, however, that the test fairly assesses the primary objectives of the lessons across curricula.

Instructional quality. Each of the four curricula had a set of lessons focused on the algebraic concept of change between two variables, and on how that change can be represented with symbolic equations. During the first year of professional development, the teachers identified those lessons in their own curriculum, and then each teacher indicated when in the course of the school year he or she would teach these lessons. Arrangements were made to observe the lessons each year and videotape either the complete set of days comprising the lessons or a sample of the days. While the project team initially strived to videotape each day of the lessons, typically 5 days, due to cost only 2 or 3 days were videotaped in subsequent years. It was assumed that instructional activities during the days observed and videotaped were representative of the type and quality for the entire set of lessons. The videotaping aimed to capture video and audio of the teacher. A total of 221 videotapes were collected and analyzed for the instructional criteria mentioned above.

Each videotape was segmented into identifiable activities, and each activity time-coded and then rated according to the five criteria (see Appendix B). All class time that was spent on the learning goal was coded and rated for IV.C: Representing Ideas Effectively. In addition, any of the other four criteria were then coded and rated for those same time segments, but these were mutually exclusive in relation to each other. For example, a segments of the lesson that focuses on the learning goal would be coded for IV.C and also, perhaps, for V.A: Encouraging Student Explanations *or* for V.B: Asking Guiding Questions, but not for both of those simultaneously.

Within each criterion, a set of three indicators were used to determine quality. Each the indicators was rated as met (score 1.0), partially met (score 0.5), and not met (score 0.0). Independent raters were contracted to conduct the evaluation. Each was trained, and a sample of videotapes was rated by multiple raters. Raters who had poor inter-rater reliability with other raters or an expert were removed and videotapes they had scored were rescored by raters with high reliability performance. Average inter-rater reliability exceeded .8 using Cronbach's reliability across samples of videotapes. Since raters varied somewhat in defining time segments, these were adjusted for this analysis so the same amount of time was examined for each rater.

Once all videotapes were coded, the average value for a videotape for each of the three criteria was computed, and an average for each teacher computed based on a time-weighted average of the criteria scores across the videotapes for that teacher. That is, the amount of time coded for a criterion was the weight function in averaging the criterion scores across the videotapes for a teacher. Therefore, the scores reflect both quantity of time spent on that criterion during a lesson and the quality of the enactment of that criterion.

Curriculum quality. All of these materials were included in the middle grades textbook analysis conducted at AAAS (AAAS, 2000). In results from that analysis, based on all available materials for the middle grades, *Connected Mathematics* and *Mathematics in Context* received the highest ratings of all materials analyzed, *Middle Grades Math Thematics* was given a moderate rating, and *Mathematics Applications and Connections* a low rating.

Professional Development. Teachers participated in three consecutive summers of professional development, but not all teachers for all three years. Some teachers participated in the study for only one year, while in the second year additional teachers participated. Two days were offered in the first summer, while 5-7 days were offered in each of the other years. In addition, a small number of sessions were given during the school year. The data set shows variability in the number of years of participation (1 to 3), as well as number of days per year, with a mean of 2.83 days and a standard deviation of 2.33.

Analysis

Since teachers varied in the number of years they participated, the missing years represent missing data. We considered only analyzing teachers with three years of data, but we concluded that too much information was lost reducing from 41 to 12 teachers. Consequently, we used multiple imputation to estimate teacher instructional scores at the teacher level. Since there was no student data for years that teachers did not participate, we imputed student scores for those years. This was done by imputing a score associated with a student that did participate in a year. The imputation was an estimate of what a student like that student would score for the year that was missing. That is, for a student with similar pre-test and post-test scores, teacher instructional quality, curriculum, number of professional development days the teacher had accumulated, and year within the three year period, how would a comparable student perform? This permitted us to estimate, with appropriate error variation, the student level performance for all three years for all 41 teachers. The amount of missing data varied from 12% to 46% for various conditions except for *Math Thematics*, in which 81% of the data were missing due to the few teachers who participated more than one year. Consequently, we analyzed results with and without that particular curriculum. The imputation literature has supported good estimation with as high as 60% missing data depending on the missing at random patterns (Schafer, 1997).

The data set follows the structure of a two-level design, with students at the first level and teachers at the second. Further, however, since teachers vary in their instructional performance from year to year, the design should also account for the cross-year dependence. This would be represented in nonzero covariances among the residuals of the fit of the second level data. Only SAS (SAS Institute Inc., 2008) permits the specification of a structured error covariance matrix at the second level using PROC MIXED. Teachers formed the second level cluster; curriculum was a 4 level categorical variable; grade was a 2 level categorical variable; and professional development days and IV.C, V.A, and V.B scores for each year were teacher-level interval covariates. Pre-test was a student-level covariate for the dependent variable post-test score. In addition, we investigated level and instructional criteria by curriculum interactions, hypothesizing that curriculum emphasis might differentiate instructional criteria quality. In preliminary analyses, none of the interactions was significant and these were not considered further.

Results

For the observed scores, we separated 7th and 8th grade students. Both groups gained on average. The 7th graders improved from 7.32 (3.90) to 9.76 (5.56), $t(1320) = 20.76$, $p < .001$, effect size of 0.57, and 8th graders improved from 9.24 (4.63) to 10.97 (5.50), $t(361) = 7.92$, $p < .001$, effect size of 0.42 . The scores were highly correlated for both grade 7 (.642) and grade 8 (.677), indicating significant relationship between initial knowledge and ending knowledge on the topics of the lesson. Means and standard deviations are shown in Table 1. Note that the means reported in the table are for all students who took either a pre-test or posttest, while the data reported above is for students who took both. It was assumed missing data within a year was random due to students missing class for reasons unrelated to the study.

In the multilevel analysis (see Table 3) we used 100 imputed data sets across the three years and covaried on the total pre-test score with a significance level of .10 to test effects because power for second level effects was not very great given the number of teachers we were able to videotape. Power was estimated as equal to about .60 for our average class size and estimated effect of 0.30 for any given predictor.

The grade effect was not significant, with a regression weight of -0.25, (s.e 0.51, $t(6979.7) = -0.49$, $p > .62$). While this may seem counterintuitive from the results above, adjusting for initial levels on the pre-test suggests 7th and 8th graders did not differ in posttest scores. In evaluating the results by year, grade was never significant. We conducted the analysis also on the unimputed original data and found the same results.

The effect of professional development was first evaluated by examining the effect of the number of days each year on that year's performance in the multilevel analysis. The regression weight was 0.08 ($t(6394) = 0.95, p > .17$). For each year, the number of days of professional development was never significantly predictive using a one-tailed test for positive effect (at $p = .10$) for any year after controlling for pre-test variation. The effect of total number of days of professional development, completed at the time of the third year's student assessment, was not a significant predictor of student achievement.

The curriculum effect on the lesson was analyzed using a contrast comparing the higher-rated textbooks (*Connected Mathematics, Mathematics in Context*) versus the lower-rated textbooks *(Math Thematics, Mathematics Applications and Connections).* In the multiple imputed data set across 10 imputations, the contrast favored the higher-rated curricula, b = 0.43, and it was significant ($t(268.4) = 1.41, p < .08$). Within each year, the curriculum contrast was not significant for year 1. For year 2, the effect was significant in the unimputed data, favoring the higher-rated curricula, $t(493) = 1.67, p < .054$, effect size = 0.12. The second year effect was not significant in the 100 imputed data sets, although in the same direction and same magnitude of regression weight, 0.42. In year 3, the effect was not significant. For each year the higher-rated textbooks had a greater effect after controlling for pre-test. Given the cross-year imputed data significance, and the similarity of estimates, we conclude the primary curriculum effect took place in year 2.

The effects of teacher instruction on the student lesson outcome were evaluated using a hierarchical model. Of the eight criteria examined across the 229 videos, only three exhibited sufficient frequency to be evaluated. The three included IV.C: Representing Ideas Effectively; V.A: Encouraging Student Explanations; and V.B: Asking Guiding Questions. Further, the authors hypothesized that the three criteria form at least a partial hierarchy in that accurate representation of ideas would be prerequisite to effective student thinking in which students would be encouraged to explain their reasoning, and to teacher guidance in that interpretation and reasoning through guiding questions. Thus, we explored the frequency and quality of accurate representations by teachers as a predictor of student achievement, followed by the quality of teachers' use of guiding questions. Preparatory to that analysis, we explored simple correlations at the teacher level.

For the multilevel analysis, we examined both the hierarchical model with IV.C entered first and followed by either V.A or V.B, and also examined the models for each separately. In the cross-year analysis, IV.C was not a significant predictor of student achievement ($t(1042) = 0.44, p > .33$), nor was V.A significant, ($t(3021.5) = .21, p > 42$). V.B was significant both controlling for IV.C ($t(1100.6) = 2.42, p < .0077$) and as a unique contributor without IV.C present ($t(1029.9) = 2.45, p < .0072$). Effect sizes in multilevel models are difficult to conceptualize, and no agreed upon method has yet been settled on. From a perspective of a simple R^2 type measure using the concept of the square of the t-statistic over t-statistic plus df, the values for the two analyses are about 1% for each. Since degrees of freedom are allocated differently under different estimation schemes, however, this is not a reliable estimate.

Examining the effect each year for category V.A, the effect was not significant in any years. For category V.B, the effect was not significant in year 1, significant for year 2 ($t(334) = 3.70$, $p < .001$, $R^2 = .11$), and significant for year 3 ($t(434) = 2.19$, $p < .015$, $R^2 = .01$). When controlling for IV.C in year 1, V.B was not significant for year 1, significant for year 2 ($t(433) = 7.55$, $p < .001$, $R^2 = .11$), and significant for year 3 ($t(1345) = 1.55$, $p < .06$). Thus, the year effects indicate the same general pattern as the multilevel cross-year model: asking guiding questions effectively promotes student learning. That the effect was present in both years 2 and 3 suggested that perhaps professional development (PD) related to asking guiding questions was a mediator of the effect. To test this, we examined the hierarchical model with professional development intervening in the model. If the V.B effect was reduced to nonsignificance and professional development was significant, the finding would support complete mediation; with both remaining significant, partial mediation.

In examining the year 2 mediation model, PD was not significant in a positive direction (it was actually significant in a negative relationship), while V.B remained significant as a positive predictor. For year 3, PD was not significant as a predictor, while V.B remained significant as a positive predictor. Thus, the PD mediation hypothesis was not supported.

Qualitative Comparative Analysis

The quantitative results to date show that the most significant aspect of instruction captured in our analysis may be teachers' ability to ask guiding questions (criterion V.B in the analysis). Another way to describe this criterion is that it measures how often and how well teachers scaffold student thinking and reasoning about the learning goals. When analysts coded for this instructional strategy, they looked for instances in which students were wrestling with some sort of mathematical dilemma. Then they examined the teacher's actions and determined the degree to which the teacher was able to respond to students' conceptual difficulties in ways that provided scaffolding towards student understanding of the learning goal. The results indicate that this strategy may have been more important for student learning than simply offering students opportunities to communicate their mathematical thinking, a strategy that showed positive effects but was not statistically significant (criterion V.A in the analysis). A careful attempt at a close comparison of selected classroom videotapes may help us gain some insights into this important finding.

Our data collection protocol was designed to enable comparisons between different teachers teaching the same lesson in a given year, between different instantiations of the same lesson by a single teacher across years, and across lessons by a given teacher within the same year. In undertaking this qualitative analysis, we sought to identify teacher pairs to serve as exploratory comparisons. It is our hope that we will be able to advance plausible conjectures about the impact of observed differences across these comparisons. This requires a careful attention to patterns of teaching as well as to nuances of teacher-student interactions that may have been noted but not fully described during the analytical coding of these lessons in the enactment of the quantitative analysis. In other words, the independent coders were trained to determine whether a sighting of a particular criterion rose to the level of partially or fully met for that criterion, but the coding process could not capture trends across multiple enactments of this criterion, nor could it capture differences between, for example, *how* two different teachers may have "fully met" a particular indicator.

To begin to provide a data-based interpretation of our findings, a close comparison has been undertaken between two teachers with similar teaching assignments but somewhat different profiles of enactment. Both teachers used materials from the *Connected Mathematics Program*, one of the highly rated materials examined in this study, and both taught first seventh, then eighth, then seventh grade mathematics over the three years of our study. This is a particularly apt comparison to examine in light of the quantitative findings in that both teachers earned above average scores for criterion V.A: Encouraging Student Explanations, with Teacher A having a higher mean score than Teacher B. However, Teacher B garnered a much higher mean score in Asking Guiding Questions (criterion V.B) and, in fact, a higher than average score across all teachers for this criterion for each of the three years of the study. Consistent with the quantitative findings, Teacher B's students showed greater than average gains across two of the three years of the study, while Teacher A's students had higher scores and gains in the first but not the final year of the study. It is probably worth noting, in light of these differential student achievement scores, that Teacher A had a class with fewer students and a higher mean SES than Teacher B. As we look closely at the teaching in these two teachers' classrooms, we are interested in characterizing differences that might shed light on this differential student achievement. We maximize the opportunity for identifying such differences by a close examination of the teaching of the *same lesson* by both teachers, with confirming evidence provided by a look at a second instance of both teachers teaching the same lesson at a second point in time.

What do we find by doing a careful analysis of the teaching in these two classrooms? We note that Teacher A routinely asks many questions of her students, both during whole class discussions and while she moves about the classroom during the exploration phase of the lesson, in an apparently purposeful way. The frequency of these questions seems to increase from the first to the third year. This is consistent with the V.A scores attributed to her teaching that increased from 0.323 in the first year to an impressive 0.595 level in her third year. Closer attention to the nature of these questions reveals a pattern that has clear implications, we believe, for student success. A graphical representation of a third year lesson of Teacher A appears below:

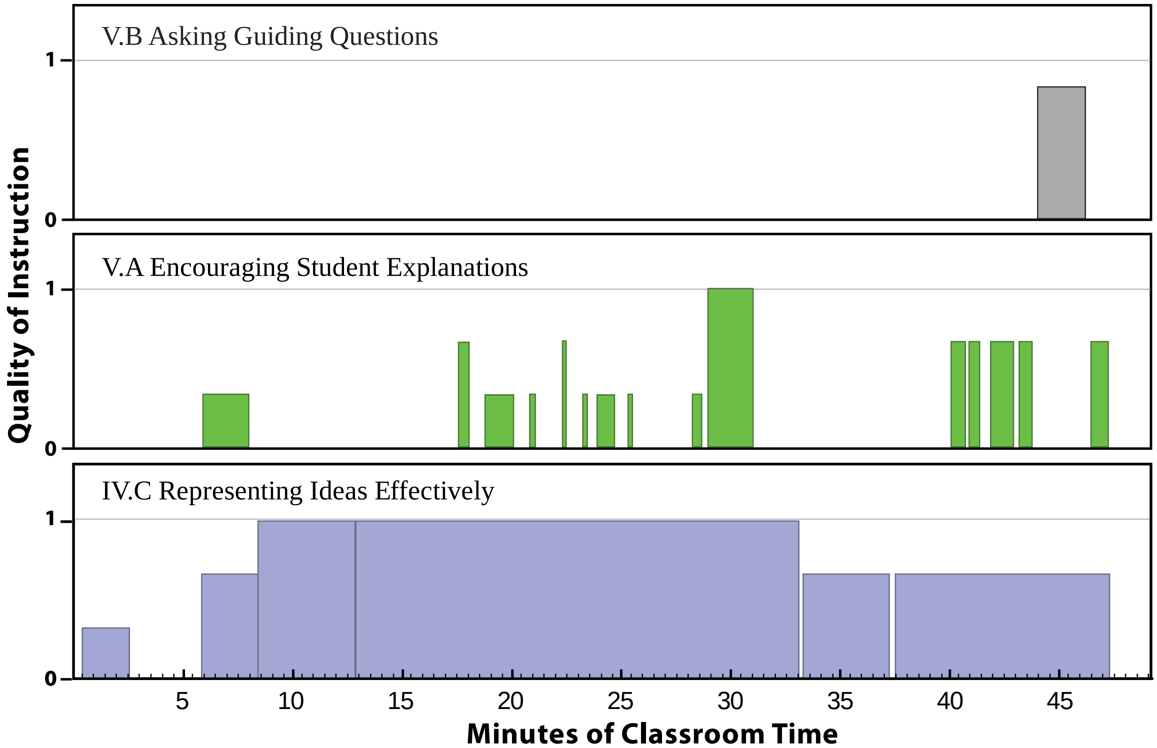


Figure 1. Multiple graphs of duration and quality of instruction according to the three criteria for Teacher A in the third year.

The blue bar represents criterion IV.C: Representing Ideas Effectively, which shows that the teacher spent most of the allotted class time on the learning goal, using representations at a moderate or high level of quality. We can see within this lesson the frequent use of Encouraging Student Explanations (criterion V.A, as represented by the green bars) and that many of these instances are brief and episodic and not altogether as fully realized as they might be given the heights of the bars. (The highest bars indicate that all three indicators under this criterion are considered achieved, the lowest bars that only the lowest of the three indicators is satisfied.) We also note only a single occurrence of the skillful use of Asking Guiding Questions, as signified by the gray bar, and this one instance occurs very near to the end of the lesson.

Upon a closer inspection of the transcript of this lesson, we find that (a) many of Teacher A's questions seem to go to matters of notational form or representational convention, and that (b) she often asks almost exactly the same carefully worded question repeatedly of different individuals during the exploration phase of the lesson. While this makes sense from the point of view of the fact that her students are working on the same mathematical task and may reflect her experience that students often exhibit the same naïve conceptions when confronted with a particular problem or problem type, the very similar nature of these questions may not directly address the thinking of the student with whom she is working at that particular moment. Put another way, questions targeted to a given misconception, however common that misconception might be, are only on target when the student in question is experiencing that particular misconception at the moment of the teacher-student interaction.

Furthermore, it seemed that Teacher A exhibited substantially more of this kind of premeditated questioning behavior in the third year of our study, perhaps in response to aspects of our professional development model. As we discussed particular lessons within the summer institutes, teachers came to appreciate the range of naïve conceptions that many students exhibited within these lessons. It may be that Teacher A took this as an incentive to hone her skill at anticipating and responding to these particular naïve conceptions or misconceptions even to the point of trying to *prevent them from occurring* with timely interventions within the lessons in question.

So, for example, within Lesson 4.1 of the 7th grade *Connected Mathematics* unit, *Variables and Patterns*, students are required to construct a table of values showing the distance traveled over several hours at an average rate of 55 miles per hour. We had discussed in our professional development the observation that many students, in filling out a sequential table of values, use an iterative additive process to complete the table rather than employing multiplicative reasoning. In other words, students are completing the table vertically by adding 55 for each entry, rather than horizontally, which would more easily translate into the symbolic equation form of distance equal to rate times time, in its explicit form. This may, in fact, be an artifact of the nature of the task, but Teacher A may have regarded it more generally as a naïve understanding to be remedied. If we examine the transcript of the videotape from her third year teaching of this lesson, we see her admonish first the entire class and then one student after another about the lure and pitfalls of the additive approach.

Here, at length, is a partial transcript of her attempt to make sure that her students see and acknowledge the multiplicative nature of rate times time as useful for computing distance. First, she addresses a pair of 7th grade male students as they begin work on Investigation 4.1 in the *Connected Mathematics* unit (second edition) of *Variables and Patterns*.

Teacher A: Okay, let's check it out. All right. How did you come up with this?

Student A: Up here it says 55 miles per hour. So I did that for distance because you're traveling that far per hour.

Teacher A: So for like 3 hours how did I, how did you come up with one hundred sixty five?

Student A: I did fifty-five times three.

Teacher A: Fifty-five times three. Okay. (Pointing to paper.) Now, does this stand for speed (noting figures in column with finger) because you have miles per hour here? What did--what do those numbers actually represent?

Student A: Miles. (Erasing.)

Teacher A: Just miles. Okay. Cause the speed is what helped you figure that out. Good.

Teacher A: Okay. (Student B), tell me what you're doing here.

Student B: I'm times-ing fifty-five times one, 55 times 2.

Teacher A: And fifty-five times . . .

Student B: Three.

Teacher A: And fifty-five times four. Okay. And what does, what do these numbers represent?

Student B: Distance.

Teacher A: Represent distance but represent distance in what?

Student B: Miles.

Teacher A: Miles? Is that on there? (Student B adds units.) Miles, okay. And you have time with hours. Okay. Thank you. Good. Come on, copy the table, let's go. Some people have already completed the graph.

Teacher A: Okay, I'm going to have the two of you talk to me at once, real quick. How did you come up with this column? What did you guys do to get this column?

Student C: We just did, umm, fifty-five each time and we kept going down.

Teacher A: Okay, you kept adding 55 each time? Okay.

Student C: But once I saw a pattern I did a hundred and ten for each other one.

Teacher A: Okay, each other was a hundred and ten. Okay. What's another pattern you see besides adding 55 each time or, every other time, adding one ten?

Student D: The one's column goes up by 5 each time.

Teacher A: Goes up by what?

Student D: By 5 each time. Oh, five zero five zero five zero. Oh, yeah, the one's column goes.

Teacher A: Okay, go beyond what you see, how can I get for five if I didn't have four, three, two and one?

Student D: Five times fifty-five.

Teacher A: What?

Student D: Five times fifty-five.

Teacher A: Five times 55. Well how do I get eight if I don't know what seven is?

Student D: Eight times fifty-five.

Teacher A: Okay. Just checking.

Although Teacher A calls on quite a few students within this fairly brief exchange, she appears to do so with one purpose in mind, to give them guided practice in moving from additive to multiplicative reasoning. Her reason for doing so may well be in response to the issues raised in our professional development about recursive to explicit reasoning. She may also be anticipating one of the culminating learning goals of this lesson, which is to write an equation to represent distance as a function of time multiplied by speed. She may well understand that generating total distance traveled by the additive process of successive summing from row to row seems inimical to or at least not supportive of this ultimate learning goal. In other words, as students determine total distance by summing distance traveled from row to row, they may not be explicitly connecting number of hours traveled to total distance traveled. Fifty-five miles are added each time to the current running total to get the successive distance traveled without reference to the total number of hours of travel. Ultimately these students are expected to write a rule expressing the relationship of distance to elapsed time and speed but this process of iteratively filling out the table may not support the development of this rule. Teacher A seems to realize this and so continues to ask students questions such as "how can I get for five if I didn't have four, three, two and one?"

Within the same lesson, Teacher B does something very different, she responds to individual students' questions and thinking rather directly. So, for example, Teacher B talks to Student E about the meaning of 55 miles per hour. Teacher B does bring a strong sense of the misconceptions that her students might make and even begins her conversation with Student E with the following observation:

Teacher B: See, (Student E), I think you guys usually think about speed in terms of how fast or slow it is, and I need you to start to think about speed in terms of how far it's going to get you. If you can keep that van at 55 miles an hour for an entire hour, do you know how far that'll get ya'?

Student E: Fifty-five miles.

Teacher B: You got it. I mean, that's the meaning of 55 miles per hour, it means you'll--you'll--you'll be fifty-five miles away from where you started every hour that you can drive like that.

But then when Student E decides to try to figure out have long it would take to travel one mile at the rate of 55 miles per hour, Teacher B attends to his redirection of the inquiry and decides to support his attempt at reasoning this out.

Student E: Fifty-five miles? That's like, that would be, you could travel a mile in like, in like . . . . . Uhhh, it's like if you're doing 55 miles an hour you can cover miles in an hour.

Teacher B: Say that one more time.

Student E: Say if, if you're traveling 55 miles an hour, then you might get . . . 55 miles in one hour.

Teacher B: Right.

Student E: Like say if you add one it's like a mile a minute.

Teacher B: It's very close to doing that. Yeah. Would you say it's more than a mile a minute or less than a mile a minute? If you're doing 55 miles in an hour?

Student E: It's probably less.

Teacher B: Well tell me this, if you could do a mile per minute now, how far would you get in an hour?

Student E: Sixty miles.

Teacher B: Right. But they're doing 55 so that means?

Student E: [Inaudible.]

Teacher B: Well, let's go back to that because you brought it up. Let's go back to how far they're going every minute. So, if they can put in sixty miles in an hour, that means they're doing a mile a minute. They're just doing 55. So, are they doing more than a mile a minute or less?

Student E: Less.

Teacher B: A little bit less.

We note a very different pattern of interaction here than that observed with Teacher A. In both cases, the teacher approaches a student engaged in an exploration with an anticipation of how that student may be thinking. Teacher A believes that many of her students will be thinking additively rather than multiplicatively as they complete the table of distance traveled. Teacher B believes that her students may have difficulty turning speed over time into distance traveled. Teacher A is very consistent in alerting her students to the multiplicative relationship of time and speed to compute distance. Teacher B anticipates that her students will not be able to convert speed into distance traveled and even gives voice to that conjecture. But while Teacher A takes her message to three different groups of students within a five minute period, Teacher B adjusts her approach almost immediately as she hears her student take a radically different approach to the problem.

Both Teacher A and Teacher B have the opportunity to respond to unexpected expressions of student thinking. For Teacher A, this occurs most obviously in the transcript when one of the students whose work she is monitoring says, "But once I saw a pattern I did a hundred and ten for each other one." Teacher A does not ignore this assertion and does understand her student's meaning but then returns rather directly to a redirection by asking, "What's another pattern you see besides adding 55 each time or, every other time, adding 110?" The second student in the pair now discovers that the one's digit alternates between 0 and 5 but the teacher is quick to assert that this is not the sort of "pattern" she is looking for, in fact her response may summarize her approach rather explicitly: "Okay, go beyond what you see, how can I get for five if I didn't have four, three, two and one?" While perhaps we should not make too much of her language here, the expression "go beyond what you see" does seem to signify to these students that she is searching for an answer that is not so easily perceived within the work itself.

Teacher B, on the other hand, is quite ready to refocus on the way her student has reframed the problem before him. When her student says, rather uncertainly, "That's like, that would be, you could travel a mile in like, in like . . ." she asks for clarification and then, when she ascertains his quest to estimate an equivalent rate in miles per minute, she restructures her questions to advance her student's agenda rather than her own. She even says, acknowledging her student's authorship of this new line of questioning, "Well, let's go back to that because you brought it up. Let's go back to how far they're going every minute."

Based upon this close reading of these two micro-interactions, we would expect the graphical representations of these two teachers' lessons to be quite different. We are not, in fact, disappointed. Recalling that Teacher A exhibited quite a few brief instances of Encouraging Student Explanations on the part of her students (the numerous green bars in the diagram of her lesson) but only a single instance of Asking Guiding Questions, we might expect more of the latter from Teacher B. Following is a graphical representation of the lesson from which the discussion with Student E is excerpted:

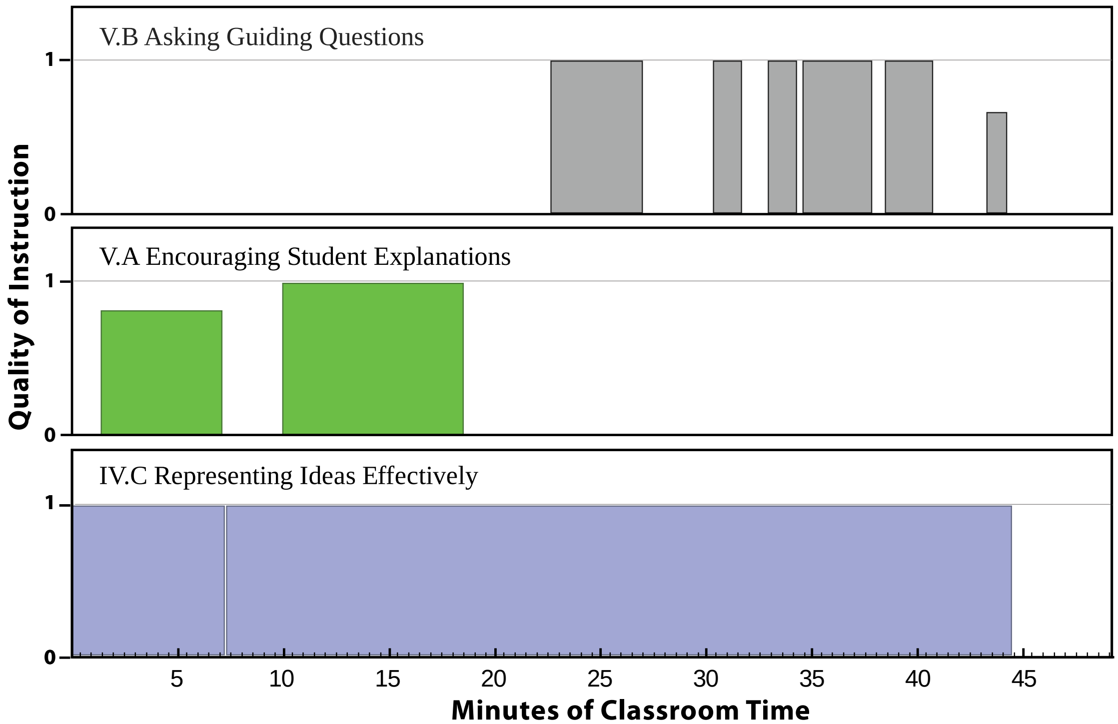


Figure 2. Multiple graphs of duration and quality of instruction according to the three criteria for Teacher B in the first year.

Teacher B's conversation with Student E occurs near the end of the day's lesson and we note that it is one of seven different interactions that were coded as representing Asking Guiding Questions. We also note that all but the last of these interactions are coded as meeting all three indicators within this criterion, including the observation that "the teacher is persistent in supporting student progress toward a deeper understanding of the learning goal(s)." Perhaps it is Teacher B's persistence in understanding and advancing her student's thinking that is most salient in the teaching-learning interaction analyzed here. While this qualitative analysis is, admittedly, largely conjectural at this point, it does seem to help us make sense of the surprises uncovered by the quantitative data analysis. Put another way, it proposes one interpretation of the story these data tell.

Our interpretation of the relationship of instructional criteria V.A and V.B to student learning might go something like this: Simply encouraging student explanations as defined in our study's criteria and indicators may not be sufficiently focused to insure that the thinking expressed advances that thinking toward the indicated learning goal. So, for example, when Teacher A asks Students B and C to identify patterns in their table of distance traveled, they would seem to assume that this is a request for them to find *numerical patterns* and proceed to offer up two candidate patterns (add 110 for every other table entry, alternate between 0 and 5 in the one's column). While these are indeed patterns, they do not advance the teacher's goal of seeing this pattern multiplicatively rather than recursively. This is not, in effect, whatever the teacher's intent, an example of Asking Guiding Questions that advance student thinking toward the desired learning goal.

On the other hand, after asking Student E, "If you can keep that van at 55 miles an hour for an entire hour, do you know how far that'll get ya'?," Teacher B listens carefully to his answer and discovers that he is wrestling with another problem, embedded in this problem but more difficult to solve than her preliminary (guiding) question. She exercises the art of guiding by beginning with the authentic question that her student poses and relating her student's inquiry to the local instantiation of the learning goal. When her student says, "Like say if you add one it's like a mile a minute," Teacher B responds with some useful feedback as well as a question that is probably intended to guide her student back to the learning goal as she has operationalized it: "It's very close to doing that. Yeah. Would you say it's more than a mile a minute or less than a mile a minute? If you're doing 55 miles in an hour?"

Perhaps the fundamental lesson in this comparison is that asking guiding questions works to advance that thinking most effectively when the teacher's questions are pointedly responsive to the actual student conceptions expressed in the moment. If, on the other hand, a teacher encourages student thinking but then does not act quite precisely to guide that thinking as expressed toward the targeted learning goal, an opportunity is lost and student learning is less likely to be advanced.

Discussion

The focus on a single learning goal across three years, almost 50 teachers in two different states, supported by four different curricula, is unique in the history of U. S. educational research. With almost 240 videos of the instruction on the algebra learning goal and assessment data on the learning goal for almost 2500 students over the three years, the resulting data set is one of the largest of its type ever collected. The analysis of teachers' instruction is also one of the most intensive ever accomplished, with the videos parsed into over 3700 separate events. While the initial results are somewhat mixed, they suggest further investigations of the data that may be productive, as we will discuss.

The finding that the amount of time spent on professional development was not a significant predictor of student learning is not greatly surprising, given the limited amount of average time spent per teacher, and the concentration of the activity in the summer. The first year of professional development consisted of only one or two days, and thus was considered "baseline." The other two years consisted of five to seven days, again in the summer. As Desimone (2009) has pointed out, other studies provide evidence that professional development is most effective when activities are spread over a semester, or when there is follow-up during the year following intensive summer work, and also when the contact time is at least 20 hours.

The curriculum effect that we found across the three years of the study is supported by some additional analyses we conducted on the materials used in this study. The original textbook analyses examined all of the materials available across multiple learning goals. We followed that up with a finer-grained analysis of the individual lessons that were selected for this study, looking for the degree to which the materials supported teachers' use of IV.C, V.A, and V.B. We found much greater support for these three instructional strategies within the highly rated materials than within the other two materials. Teachers using the other two materials, even with low fidelity, were operating in a context of weak support for encouraging student explanations or for the use of guiding questions.

A limitation of the research in this study to date is our focus only on the total algebra achievement score. With 16 items representing three major constructs, we expect further research to refine our results here and perhaps bring greater clarity to the high variability we have seen so far in this extensive study. The extended response item that directly targets the learning goal can be one aspect to consider, and there are clusters of items that assess students' understanding and use of representations, while other clusters assess their ability to explain their mathematical ideas.

In addition, we have further data, not reported here, on a second learning goal in the content area of Data Analysis. For that learning goal we have additional instructional data (in the form of analyzed videos) and student achievement data (in the form of a second set of student assessments) for the same set of teachers as we report on here. That additional data may help shed further light on the relationships we have seen thus far for the algebra learning goal. In particular, we will be able to determine whether the teachers' use of guiding questions has the same significant effect for the 6th grade teachers on a Number learning goal, and whether it holds for all three grades on a learning goal that was not the content focus of the professional development.

Table 1: Means and Standard Deviations of Student Algebra Scores

		Algebra Pre-test			Algebra Post-test		
Variable	Condition	N	Mean	SD	N	Mean	SD
Algebra	Overall	2189	7.59	4.294	1975	9.91	5.51
	Year 1	956	6.43	3.62	816	8.43	VB5
	Year 2	710	8.83	4.83	644	10.48	VB5
	Year 3	523	8.05	4.11	515	11.56	5.64
	Grade 7	1738	7.26	4.16	451	8.88	4.55
Curriculum	CMP	667	7.12	3.81	637	10.13	VA3
	Math App	460	7.08	3.80	411	8.77	5.46
	MthM	355	6.91	4.02	243	9.20	5.41
	MIC	707	8.72	4.91	684	10.66	5.77
High vs. Low Rated	Low	815	7.01	3.90	654	8.93	5.44
Curricula	High	1374	7.94	3.90	1321	10.40	5.48

Table 2: Teacher-level Variable Means and Standard Deviations

Variable	N	Mean	SD			
Professional Development						
Overall	41	5.66	4.69			
Year 1	41	0.58	0.87			
Year 2	41	3.08	2.51			
Year 3	41	2.50	2.85			
Category						
IV.C (Representing Ideas Effectively)				V.A (Encouraging Student Explanations)		
	N	Mean	SD	N	Mean	SD
Year 1	20	0.78	0.11	20	0.27	0.12
Year 2	25	0.75	0.16	25	0.21	0.13
Year 3	20	0.76	0.15	20	0.29	0.18
V.B (Asking Guiding Questions)						
Year 1	20	0.10	0.64			
Year 2	25	0.13	0.12			
Year 3	20	0.11	0.08			

Table 3: Multiply Imputed Cross-Year Multilevel Analysis of Algebra Achievement

Parameter	Estimate	Std Error	95% Confidence Limits		DF	Minimum	Maximum		t for H0: Parameter= Theta0	Pr > |t|
Intercept	5.277581	3.78903	-2.15073	12.70589	4610.7	1.219247	7.998252	0	1.39	0.1637
PRTOT	0.779166	0.028607	0.72307	0.83526	2720.1	0.749772	0.812711	0	27.24	<.0001
Grade	-0.33328	0.488904	-1.29164	0.62508	9213.6	-0.72491	0.063559	0	-0.68	0.4954
PD	0.082427	0.087152	-0.08842	0.25327	6394	0.000158	0.162812	0	0.95	0.3443
hilo	0.360377	0.248727	-0.12724	0.848	4681.1	0.111844	0.585831	0	1.45	0.0737*
IVC	0.300726	1.422364	-2.49047	3.09193	988.53	-1.5797	2.245671	0	0.21	0.8326
VB	6.07455	2.507714	1.15411	10.99499	1100.6	3.54246	9.793233	0	2.42	0.0156

Note: PRTOT = pre-test Algebra; PD = professional development days; hilo = High vs. Lower rated curricula

References

American Association for the Advancement of Science. (2000). Middle grade mathematics texts: A Benchmarks-based Evaluation. Accessed at http://www.project2061.org/publications/textbook/mgmth/report/default.htm.

Billstein, R., Williamson, J., Montoya, P., Lowery, J., Williams, D., Buck, M., Burkett, C., Churchill, L., Clouse, C., Denny, R., Derrick, W., Dolezal, S., Galarus, D., Kennedy, P., Lamphere, P., Merrill, N., Morse, S., Petit., M., Runkel, P., Sanders-Garrett, T., Seitz, R., Spence, B., Sowders, B., Tuckerman, C., Wenger, K., Wilkie, J., Wilson, C., & Winston, B. (1999). *Middle grades math thematics*. Evanston, IL: McDougal Littell.

Collins, W., Dritsas, L., Frey-Mason, P., Howard, A. C., McClain, K., Molina, D. D., Moore-Harris, B., Ott, J., Pelfrey, R. S., Price, J., Smith, B., & Wilson, P. S. (1998). *Mathematics applications and connections*. Columbus, OH: Glencoe/McGraw-Hill.

Collis, K., Romberg, T., & Jurdak, M. (1986). A technique for assessing mathematical problem-solving ability. *Journal for Research in Mathematics Education, 17*(1), 206-221.

Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199.

Kaput, J. (1999). Teaching and learning a new algebra. In E. Fennema & T. Romberg (Eds.), *Mathematics classrooms that promote understanding* (pp. 133-155). Mahwah, NJ: Erlbaum.

Kieran, C. (n.d.). What do students struggle with when first introduced to algebra symbols? *National Council of Teachers of Mathematics Research Brief*. Retrieved January 16, 2008 from http://www.nctm.org/news/content.aspx?id=12332.

Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Phillips, E. P. (1998). *Connected mathematics*. Menlo Park, CA: Dale Seymour.

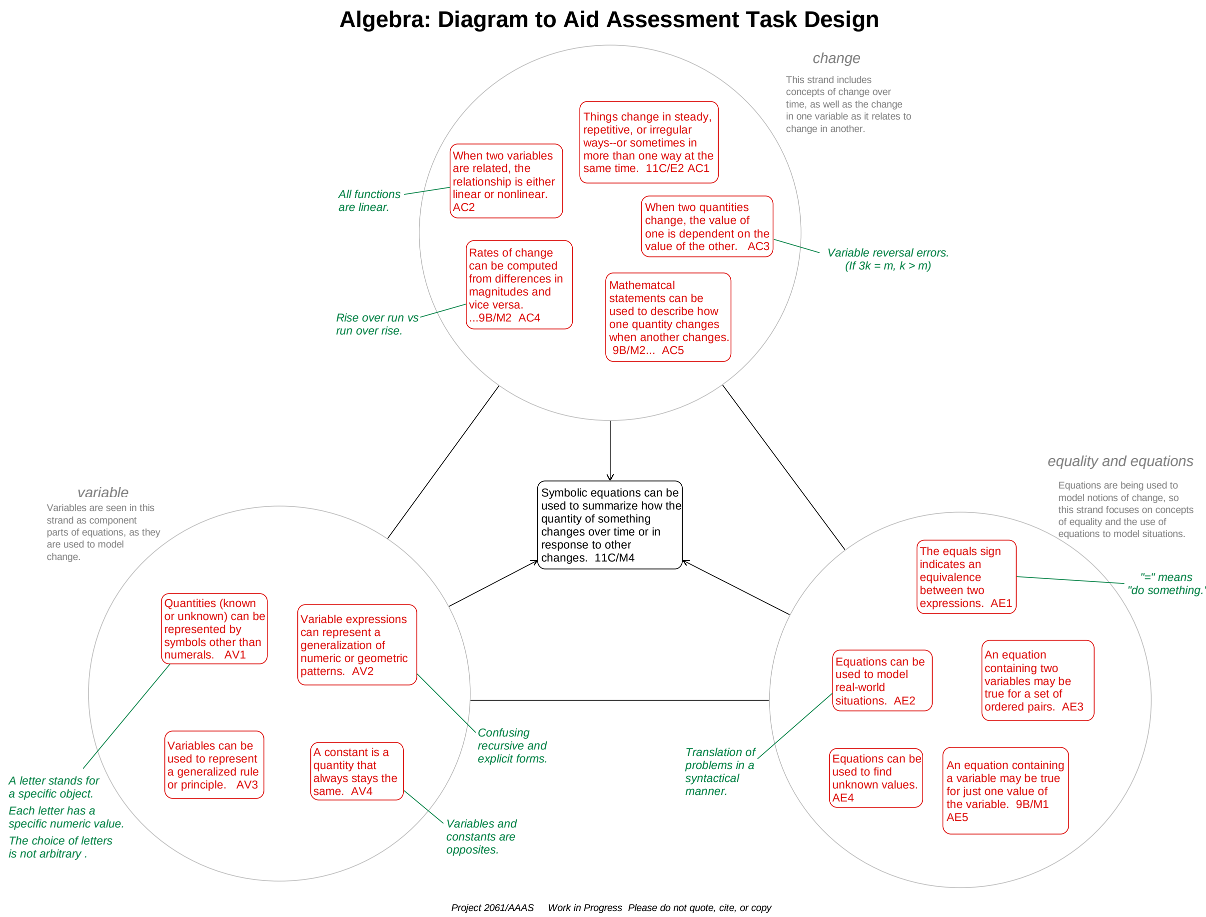
Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Phillips, E. P. (2000). *Connected mathematics*. Glenview, IL: Prentice Hall.

Romberg, T., Burrill, G., Fix, M., Middleton, J., Meyer, M., Pligge, M., Brendefur, J., Brinker, L., Browne, J., Burrill, J., Byrd., R., Christiansen, P., Clarke, B., Clarke, D., Cole, B., Dremock, F., Halevi., T., Milinkovic, J., Shafer., M., Shew., J., Schultz, K., Simon, A., Smith, M., Smith, S., Spence, M., & Steele, K. (1998). *Mathematics in context.* Encyclopedia Britannica Educational Corporation.

SAS Institute Inc. (2008). *SAS/STAT software, Version 9.2 of the SAS System for Windows*. Cary, NC: Author.

Schafer, J. L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.

Appendix A


9/30/02
Algebra: Diagram to Aid Assessment Task Design
change
This strand includes concepts of change over time, as well as the change in one variable as it relates to change in another.
Things change in steady, repetitive, or irregular ways--or sometimes in more than one way at the same time. 11C/E2 AC1
When two variables are related, the relationship is either linear or nonlinear. AC2
All functions are linear.
When two quantities change, the value of one is dependent on the value of the other. AC3
Variable reversal errors. (If 3k = m, k > m)
Rates of change can be computed from differences in magnitudes and vice versa. ...9B/M2 AC4
Rise over run vs run over rise.
Mathematcal statements can be used to describe how one quantity changes when another changes. 9B/M2... AC5
Symbolic equations can be used to summarize how the quantity of something changes over time or in response to other changes. 11C/M4
variable
Variables are seen in this strand as component parts of equations, as they are used to model change.
Quantities (known or unknown) can be represented by symbols other than numerals. AV1
Variable expressions can represent a generalization of numeric or geometric patterns. AV2
Variables can be used to represent a generalized rule or principle. AV3
A constant is a quantity that always stays the same. AV4
A letter stands for a specific object.
Each letter has a specific numeric value.
The choice of letters is not arbitrary .
Confusing recursive and explicit forms.
Variables and constants are opposites.
equality and equations
Equations are being used to model notions of change, so this strand focuses on concepts of equality and the use of equations to model situations.
The equals sign indicates an equivalence between two expressions. AE1
"=" means "do something."
Equations can be used to model real-world situations. AE2
An equation containing two variables may be true for a set of ordered pairs. AE3
Equations can be used to find unknown values. AE4
An equation containing a variable may be true for just one value of the variable. 9B/M1 AE5
Translation of problems in a syntactical manner.
Project 2061/AAAS Work in Progress Please do not quote, cite, or copy

Appendix B

Criteria Used for Measures of Teaching Practice

IV.C: Representing Ideas Effectively. Does the teaching include accurate and comprehensible representations of the learning goals?

Indicators for meeting Criterion IV.C:

1. The teaching accurately represents relevant aspects of the learning goal(s), or brings out the limitations of a representation(s) or model(s) that is not accurate. In the case of student-generated representations, the teaching points out inaccuracies in those representations.

2. The representation(s) or model(s) is likely to be accessible and/or familiar to students.

3. The teaching effectively uses the representation or set of representations to promote a conceptual understanding of the learning goal.

V.A: Encouraging Student Explanations. Does the teaching routinely encourage each student to express, clarify, justify, interpret, and represent his/her ideas [knowledge/understanding] about the learning goals (e.g. with tasks, real world examples, representations, and/or readings related to the learning goals)?

Indicators for meeting Criterion V.A:

1. The teaching encourages students to express their knowledge/understanding relevant to the learning goals.

2. The teaching encourages students not only to express but also to clarify, justify, interpret, and/or represent their knowledge/understanding.

3. The teaching provides opportunities for each student (rather than just some students) to clarify, justify, interpret, and/or represent their knowledge/understanding.

V.B: Asking Guiding Questions. Does the teaching include questions and/or tasks that guide student interpretation and reasoning about the learning goal(s)?

Indicators for meeting Criterion V.B:

1. The teaching uses specific questions and/or specific tasks to address a **mathematical dilemma** that confronts the student(s) and to support student progress toward a more complete conceptual understanding of the learning goal(s) without leading.

2. The guiding questions/tasks are responsive to evidence of student thinking rather than generic in nature and **directly target** the student's mathematical dilemma regarding the learning goal(s).

3. The teacher is **persistent** in supporting student progress toward a deeper understanding of the learning goal(s).

II.C: Finding Out Students' Ideas. Does the teaching use questions and/or tasks to identify what their students think about familiar situations and/or phenomena related to the benchmark ideas before these goals are introduced?

Indicators for meeting Criterion II.C:

1. The teaching poses specific questions and/or tasks to identify students' ideas related to the benchmark ideas before the ideas are introduced.

2. The questions and/or tasks go beyond identifying students' knowledge of terms, such as asking students to make predictions and/or give explanations of phenomena.

3. The questions and/or tasks are likely to be comprehensible to students who are not familiar with the benchmark ideas introduced in the unit.

4. Indicator not observable.

5. The teaching probes beneath students' initial responses to questions and/or tasks using follow-up questions and/or tasks.

6. Indicator not observable.

VI.B: End of Lesson Assessment. Does the teaching engage students in assessment questions and/or tasks that require students to show, use, apply, explain, or otherwise demonstrate their understanding of the knowledge and/or skills specified in the learning goals?

Indicators for meeting Criterion VI.B:

1. The teaching engages students in assessment questions and/or tasks that focus on the use of the knowledge and/or skills specified in the learning goals (for example, they ask students to apply, explain, etc.).

2. The teaching engages students in questions and/or tasks that students have not previously encountered.

3. The teaching engages students in questions and/or tasks that do not use the exact context in which students were taught the learning goals.

4. The teaching engages students in assessment questions and/or tasks likely to provide evidence about whether or not students hold any commonly held student ideas that are relevant to the learning goals and described in the learning research literature.