FACTORS IN THE ACHIEVEMENT OF PRESERVICE ELEMENTARY TEACHERS IN MATHEMATICS CLASSES

Raven McCrory
Michigan State University
mccrory@msu.edu

Changhui Zhang
Michigan State University
zhangcha@msu.edu

Andrea Ploucher Francis
Michigan State University
ploucher@msu.edu

Sarah Young
Michigan State University
youngsa1@msu.edu

While it has been found that teacher knowledge affects mathematics student achievement, to date, little research has explored how professors affect preservice elementary teacher's mathematical knowledge. This study explores future teachers' learning in undergraduate mathematics classes. Our data includes pre and post tests from over 1000 students in classes of 41 instructors at 17 institutions in four states as well as teacher survey data from those 41 instructors. In our multilevel models, we identify three key variables that influence gain: students' prior knowledge, use of specifically designed textbooks, and the methods the professor uses to teach the course.

Introduction

Improving the mathematical knowledge of elementary teachers is key to improving children's mathematical knowledge (e.g., Conference Board of the Mathematical Sciences, 2001; Hill, Rowan, & Ball, 2005; Monk, 1994; Rowan, Correnti & Miller, 2002). Mathematics classes required for certification provide a unique opportunity to influence teachers' mathematical knowledge, yet there has been little research about either what is offered or what students learn in these undergraduate mathematics classes. In this study, we begin to address these questions.

On average, undergraduate programs require teachers to take 2.7 mathematics courses for elementary certification and in states where there is a separate endorsement or certificate, 5.6 classes for middle school certification, up from 2.4 and 3 respectively in 2000 (Lutzer, Rodi, Kirkman, & Maxwell, 2007). Although the number of mathematics classes required of these future teachers has been increasing over the last decade, we know almost nothing about the content of these courses, who teaches them, or what their impact is on future teachers' mathematical knowledge. The goal of our research is to understand what these future teachers learn and what accounts for differences in learning across instructors at various institutions. Specifically, we ask what characteristics of students, courses, instructors, and institutions explain variation in achievement across mathematics classes for future teachers that are focused on number and operation?

This work is important because these undergraduate classes are a unique sustained opportunity to influence teachers' mathematical knowledge. If we could learn more about how to make these classes better – to have a greater impact on teachers' knowledge – through changing the content, the textbook, the way classes are taught, or other variables, our findings could have an enormous impact on the preparation of elementary teachers for teaching mathematics

Background

To investigate this question, we designed a multilevel study at undergraduate institutions in 4 states. We collected data from future teachers (students in undergraduate mathematics classes), instructors, and mathematics department chairs. We developed hypotheses about what variables might predict student learning and tested these hypotheses using multilevel analyses. The complete study is described in detail in other documents including McCrory (2009).

Literature Review

The National Math Panel (NMP) report, (see p. 5-1) which summarizes research on teacher knowledge, suggests that the effect of teacher quality on student achievement is large. In one study, 12-14% of the variation in student achievement was attributed to differences in teachers. In another study, the difference in outcomes for students of the worst and best teachers was 10 percentile points on a mathematics assessment. Yet, the NMP report points out that understanding the individual differences that constitute teacher quality remains elusive. One likely candidate is teacher knowledge of mathematics, but there, the research is inconclusive.

Studies of teacher knowledge have typically relied on proxy data (number of math classes, certification status, years of experience, test scores such as SAT or ACT) to investigate the question of what mathematics teachers bring to their teaching. One exception is the work of Ball and colleagues at the University of Michigan (e.g., Hill, Rowan, & Ball, 2005; Hill Schilling & Ball, 2004). In their project, Learning Mathematics for Teaching (LMT), they developed measures of elementary teachers' mathematical knowledge and went on to show that the knowledge measured by their instrument was a significant predictor of K-8 student achievement in mathematics. That is, teachers who scored higher on the LMT measures had students who scored higher on standardized mathematics achievement tests. Their work shows that there is specific mathematical knowledge that contributes to or is an indicator of teacher quality.

What we do not know is whether or how prospective teachers might learn this content. It seems unlikely that they learn it from their high school mathematics classes or from conventional college mathematics courses. If that were the case, the problem of elementary teachers' mathematical knowledge would be nonexistent or at least easy to address. Although some have argued that more required mathematics classes would improve teacher quality, research suggests otherwise: taking more mathematics courses does not necessarily result in teachers who teach mathematics more effectively. Wilson, Floden and Ferrini-Mundy (2002) point out in their report on teacher preparation that studies about subject matter preparation "undermine the certainty often expressed about the strong link between college study of a subject matter and teacher quality" (p. 191). What is missing is a better understanding of the content that matters and how best to offer it to future teachers.

Purpose of this Study

This study begins to address what affects prospective teachers' mathematical content knowledge in their mathematics courses. In particular we investigated a number of factors that we hypothesized might influence their mathematical achievement, measured with the LMT items that we know correlate with teacher quality. Factors include the textbook used, the instructor's attitude toward the class and experience teaching the class; and how the textbook was used; and methods of teaching. Other factors, related to the students themselves, are prior knowledge, socio-economic status, and attitude toward mathematics.

Method

Population

Data for this study were collected at institutions in 4 states, chosen to reflect variation in state policy and K-8 mathematics outcomes. The analysis here includes data from 41 instructors and 1706 students at 17 institutions. The data reported in this paper were collected between September 2006 and December 2008.

Instruments and Data Collection

Students completed pre- and post-tests using items from the LMT project. The pre and post-tests were different, equated through Item Response Theory (IRT) methods to make results comparable. We used two forms and 6 additional common items. Each student took one of the forms plus the 6 common items for the pretest and the other form for the posttest, making the pretest somewhat longer than the posttest. Thus, every student completed every item, but had completely different pre- and post-tests. The student tests also included attitudes and beliefs items and demographic questions as explained below. The pretest was administered in the first two weeks of class; the posttest in the last two weeks. LMT items are not generally available for public release, but a set of released items is available on the LMT Web site, http://sitemaker.umich.edu/lmt/measures.

Instructors completed an extensive survey developed for this project. It includes questions about content coverage, teaching methods, contextual issues, personal demographics, and more. The complete instrument is available at our Web site, http://meet.edu.msu.edu. The instructor survey was administered at the end of the semester during which student pre/post tests were administered.

Data Analysis and Results.

The first analyses were on the test scores themselves. We used IRT parameters from the University of Michigan project to calculate pre- and posttest scores for all students. IRT scoring takes into account the difficulty of items and thus makes it possible to compare pre- and posttest scores on the same scale. The scores, however, do not correspond to percent correct and are more like z-scores. For reporting purposes, we set the average pretest IRT score to 50 with a standard deviation of 10. The range of scores is theoretically infinite, but practically, scores fall within 3 standard deviations of the average. Posttest scores are calculated using the same parameters and are placed on the same scale as the pretest. In these data, the pretest mean is 50.00, with a posttest mean of 59.16.

Table 1: Student (Future Teacher) Descriptive Data

Variables	Coding and Range	Mean	SD
Pretest Score	17 – 79	50.00	10.00
Prior Knowledge (CACT = SAT or ACT on a common scale)	12 – 36	23.17	4.38
I like Math	0 = Strongly disagree, disagree, undecided 1= Strongly agree or agree (Used in model)	0.46	0.50
	On 5 point scale (used in correlations)	3.03	1.26
College Math Coursework	0 = none 1 = 1 2 = 2 3 = 3 4 = 4 or more	2.47	1.12
SES (Mother's Education)	0 = no higher ed, 1 = some higher ed	0.46	0.50

Descriptive data for students and instructors are shown in Tables 1, 2 & 3. We developed a measure of prior knowledge for students using self-reported ACT or SAT scores. We put these on a common scale (1-36) using conversions published by ETS and named that variable CACT.

We used a single question from the student survey to measure their attitude toward mathematics. They ranked from 1 (Strongly disagree) to 5(Strongly agree) the statement “I like mathematics”. We converted their responses to a two-point scale: 0 for a response of 1, 2 or 3; 1 for a response of 4 or 5.

Table 2: Correlations of Student (Future Teacher) Descriptive Data

	1	2	3	4	5	Variable Explanation
1	1.00	**0.49	**0.48	-0.02	**0.20	Pretest Score
2		1.00	**0.41	-0.33	**0.21	Post test Score
3			1.00	-0.01	**0.31	Prior Knowledge (C_ACT)
4				1.00	**0.13	College Math Coursework
5					1.00	“I like math” (5 point scale)

*Correlation significant at the .01 level, two- tailed.

For instructors, we asked them what textbook they used for the course. Based on their response and on the list of 13 textbooks in print written specifically for such a course (list available on the Web site http://meet.educ.msu.edu), we created a variable that had the value 1 if they use one of the textbooks on the list as their primary textbook, 0 if they use some other textbook or do not use a textbook at all. We also asked them about their interest in teaching the course before the current semester, and their interest in teaching the course again. Results of these questions are shown in Table 3, along with information about class size, and experience.

Table 3: Instructor Descriptive Data

Variables	Coding and Range	Mean	SD
Primary Textbook from choice of 13	1 = a primary textbook on our list 0 = not a textbook on our list	0.65	
Class Size	4 – 102	26	13
CACT (mean SAT or ACT on a common scale)	20 – 27	23	1.6
Years College Teaching Experience	0 – 41	16	11
Interest in teaching this course	0=no interest at all 1=limited interest 2=some interest 3=a great deal of interest	1.26	0.59
Interest in teaching this course again	0=no interest at all 1=limited interest 2=some interest 3=a great deal of interest	1.22	0.55
Teaching Methods (Mean of 11 items)	1.45 – 4.00	2.74	0.58

Another variable we developed was a measure of teaching methods. For this variable we asked instructors: “In your mathematics course, how often do your students engage in each of the following activities? Please check the box that best describes what happens in your course.” The scale was from 1 to 4: 1. Never or almost never; 2. Some lessons; 3. Most lessons; 4. Every lesson. The 11 items instructors ranked included (complete list available in McCrory, 2009):

- Explain the reasoning behind an idea
- Work on problems for which there is no immediate method of solution
- Listen to you explain terms, definitions, or mathematical ideas (Reversed)
- Listen to you explain computational procedures or methods (Reversed)

Since these are questions about what the instructor expects *students* to do, the last two on the list were reverse coded to create a scale that indicates student's personal engagement with the mathematics as compared to listening to the instructor. At one extreme (4), students would be doing mathematics at all times. At the other extreme (1), students would be listening to the instructor at all times. As table 3 shows, the mean score on this variable was 1.81 suggesting that these instructors use a mixture of methods, leaning slightly toward student engagement. Although the dataset includes many more variables than described here, we include only those used in the models developed thus far.

To put the student and instructor data together and investigate our hypotheses, we developed a growth model using HLM (Raudenbush & Bryk, 2002). Although growth models are most often used with more than two data points to measure growth, we chose this model because it allowed us to interpret the data more completely than a two level model with either gain or posttest score as outcome, and it allowed us to use level 1 data from students who took only the pre or post test. Because we are using scores from an Item Response Theory (IRT) model, it is possible to estimate the growth model with only two data points. In this model, we define three levels. Level 1 is the growth level with time set to 0 for the pretest, 1 for the posttest. Level 2 is the student level. Level 3 is the instructor level. We do not have adequate data for an institutional level, and have not developed a state-level model (which would include only 2 states because of the sparse data in the other 2 states).

Figure 1: Unconditional model for allocating variance

Level 1- Growth:

$$Y = P0 + P1*(TIME) + E$$

(TIME is 0 or 1)

Level 2 - Student:

$$P0 = B00 + R0$$
$$P1 = B10 + R1$$

Level 3 - Instructor:

$$B00 = G000 + U00$$
$$B10 = G100 + U10$$

(E, R, and U are random error)

The unconditional model is used to allocate variance. There are no predictors in the model and what we learn from it is how much of the variance is within instructor and how much between instructors. Although we are interested generally in explaining achievement, the primary purpose of this analysis is to explain variation between instructors. This model is shown in Figure 1.

From this model we get the following data:

Mean Pretest score (B00): 50.56
Average gain (B10): 7.73
Gain Score student level variance: 3.72
Gain Score Instructor level variance: 7.80

Note that the mean score of 50.56 differs from the set mean of 50 because even in this simple model, Bayesian estimation techniques are used and error is in play. For our models, we assume that there is measurement error, and we use expected a posteriori (EAP) procedure for calculation of parameters. Variance is a pure number without units, and this tells us that if we assume there is measurement error, we have a lot of variation between instructors relative to the variance within classes.

We built a number of models to test hypotheses and found no significance for student SES, instructor experience, instructor attitude toward the class, or class size. Significant predictors

Figure 2: Model 2 with preditors at student and instructor levels

```
Level 1: Growth Model
        Y = P0 + P1*(TIME) + E
         (Time is 0 or 1)
Level-2: Student Level with CACT
        P0 = B00 + R0
        P1 = B10 + B11*(CACT) + R1
Level-3: Instructor Level with Textbook and Methods
        B00 = G000 + U00
        B10 = G100 + G101(TEXTBOOK) + G102(METHODS) + U10
        B11 = G110
```

include student CACT, student attitude toward mathematics, textbook used, and method of instruction. We found that CACT and student attitude were correlated, so that if both are in the model, one loses significance. We chose to use CACT rather than student attitude in our models because of the measure's greater reliability. Although we have not completely tested all hypotheses, the model that best predicts student outcomes so far is shown in Figure 2.

Results of Model 2, shown in Table 4, are:

Mean Pretest: 51.48
Average Gain, no primary textbook, average methods: 4.79
Average Gain, Add primary textbook 4.58
Average Gain, Change in methods by one point: 2.92
Average Gain, Change in CACT by one point: 0.52

Again, the mean pretest is different from 50 and also different from the unconditional model because the estimation method is recursively using the data to come up with the most likely "true" value of the mean pretest. In this model, we predict that an instructor with average methods score of 2.73 (see Table 3) who did not use one of the 13 textbooks on our list would have an average posttest score in his/her class of 51.48 + 4.79 = 56.27. If the instructor used one of the textbooks, the predicted posttest score would increase by 4.58 points; and if the instructor scored a point higher on the methods measure, the predicted score would increase by 2.92.

There are numerous issues with these models and with the results, not least of which are the set of decisions about the role of error; which parameters are fixed and which are allowed to vary by instructor (e.g., in Model 2, we did not let CACT vary by instructor, indicating an assumption that instructors are equitable with respect to prior knowledge); and which IRT parameters to use. We have tried the models with different assumptions, and although the absolute numbers are different, we consistently see that the use of one of the textbooks and the use of different teaching methods are significant – statistically and practically – at the instructor level; and the CACT score (a measure of prior knowledge) is significant at the student level. The more prepared students are for this class (as measured by SAT or ACT), the better they do on the tests. That is no surprise, but it may be a surprise that the CACT and attitude toward mathematics are equally good predictors (raising a chicken and egg question that cannot be answered with our data). We also checked the assumption that instructors are equitable with respect to CACT and found it to be correct.

Table 4: Model 2 Results

	Coefficient	Error	T-Ratio	d.f.	P-Value
For INTRCPT1, P0 (pretest score)					
For INTRCPT2, B00					
INTRCPT3, G00	51.48	0.60	85.82	37	0.000
For POST slope, P1 (Gain score)					
For INTRCPT2, B10					
INTRCPT3, G100	4.79	1.08	4.42	35	0.000
TEXTBOOK, G101	4.58	1.35	3.40	35	0.002
METHODS, G102	2.92	1.10	2.70	35	0.011
For CACT, B11					
INTRCPT3, G110	0.53	0.07	7.73	952	0.000

Discussion

Several results stand out in this analysis. First, students' attitudes toward mathematics are very important. In particular, whether they like math or see themselves as capable of doing mathematics are both significant predictors of their achievement. This is no surprise given the history of research on attitudes toward mathematics, but it signals a possible leverage point for improving teacher knowledge. If we could make inroads into improving future teachers' confidence in their mathematical ability or their attitude about mathematics, we might be able to teach them more effectively. An alternative hypothesis is that they dislike mathematics because they are bad at it and that their self-assessment is accurate. This depressing view is not borne out by research that suggests that students' mathematical abilities are not fixed, but depend on their effort and engagement.

Second, students' prior knowledge is important. If students enter the class with a weak background, these classes generally do not make up for these deficits, although some instructors are more successful here than others. On the encouraging side, even students with weak prior knowledge are, on average, learning from these classes.

Third, students are learning a lot from these classes. A gain of 9 points is almost a full standard deviation. This is a big gain by any standards, and since we have reason (from LMT) to believe that what they are learning matters in their future teaching, this is an important finding.

Fourth, students in classes of instructors who use one of the textbooks designed for a mathematics class for future elementary teachers are more successful than those in classes that do not use such a book. This may be because the books tend to treat the mathematics more coherently or appropriately than self-developed materials or other mathematics textbooks not intended for such classes. Alternatively, it may be that the textbook signifies something else about the class that matters. For example, it could be that instructors who use one of these textbooks teach in a more consistent manner across the semester, or that the department syllabus based on the textbook is better developed. There are many other possibilities that could explain this result. Without more data from different textbooks, there is little we can say to explain this further.

Finally, and perhaps most surprisingly, the teaching method matters. Teachers who report more student engagement with mathematics and less lecturing tend to have greater student achievement. This may be related to the first point in this discussion: students do better when they have a more positive attitude toward mathematics. A class in which they have an opportunity to work directly on mathematics rather than only listen to mathematics may be instrumental in developing a more positive attitude. One could argue that agreeing to participate in the ME.ET project signals a different set of beliefs or commitments to mathematics education, making the method result an artifact of the particular instructors involved in this study. But we did have considerable variation on this measure, suggesting that even if these instructors are different than nonresponders, they are also different from one another.

We are still analyzing these data, and will have a more complete report on model development. We will also develop structural equation models to try to tease out more complex relationships among variables (e.g., that the teaching methods may result in more positive attitudes that then impacts learning). In the meantime, our results suggest that there are leverage points in the courses that could be used to improve the mathematical knowledge of future elementary teachers.

Acknowledgements

This research is funded by the National Science Foundation (Grant No. 0447611). The authors wish to thank the instructor who generously participated in this project and the other team members – Ga Young Ahn, Rachel Ayieko, Marisa Cannata, Beste Güçler, Rae-Young Kim, Young-Yee Kim, Jane-Jane Lo, Jessica Liu, Jungeun Park, and Helen Siedel – who collected data and participated in discussions that made our analysis possible.

References

Conference Board of Mathematical Sciences (2001). *The mathematical education of teachers*. Washington, DC: American Mathematical Society.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371-406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11-30.

Lutzer, D. J., Rodi, S. B., Kirkman, E. E., & Maxwell, J. W. (2007). *CBMS 2005 Survey: Statistical Abstract of Undergraduate Programs in the Mathematical Sciences in the United States*. Washington, DC: American Mathematical Society, Conference Board of the Mathematical Sciences.

McCrory, R. (2009). *ME.ET progress report*, available online at http://meet.educ.msu.edu/reports/

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125-145.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical Linear Models* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Rowan, B., Correnti, R., & Miller, R. J. (2002). *What large-scale survey research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools*. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education, 53*(3), 190-204.