The Impact of Inquiry-Based Mathematics Courses on Content Knowledge and Classroom Practice

Rachel Cochran, Center for Educational Accountability

John Mayer, University of Alabama at Birmingham

Bernadette Mullins, Birmingham-Southern College

Abstract: This study includes middle grades teachers participating in the first of a series of inquiry-based mathematics content courses as well as pre-service teachers enrolled in an inquiry-based university course. A variety of data sources (including objective assessment, performance assessment, portfolios, a behavioral checklist, classroom observations and participant surveys) are used to provide a comprehensive picture of participants as learners and teachers of mathematics. This paper describes changes in participants' content knowledge and classroom practice.

1. Introduction

The focus of this study is on middle grades teachers participating in the first of a series of summer mathematics content courses as well as pre-service teachers enrolled in a revised university course. These courses were offered as part of the Greater Birmingham Mathematics Partnership (GBMP), a Math Science Partnership funded by the National Science Foundation involving nine school districts, two institutes of higher education (IHE) and a nonprofit organization. GBMP seeks to involve all stakeholders in improving mathematics education and therefore partnership activities include working with parents and members of the public, community and business leaders, superintendents and school administrators, in-service teachers, pre-service teachers, and IHE faculty. This study focuses on the impact of the content courses for in-service and pre-service teachers.

The course *Patterns, Functions, and Algebraic Reasoning*, developed by the Mathematics Education Collaborative, is an intensive nine-day summer mathematics content course. It is part of an innovative approach to professional development in mathematics education in which teachers are offered opportunities to struggle with complex, rich, and expandable mathematical tasks. Collections of tasks are chosen to help participants develop a deep understanding of some of the fundamental ideas in mathematics and build confidence in mathematical problem-solving. During the course, teachers are engaged in inquiry-based mathematics investigations in both group and individual settings. A variety of manipulatives (color tiles, pattern blocks, interlocking cubes, Cuisenaire rods) are available to aid visualization of patterns. The instruction in the course was designed to further teachers' conceptual understanding, procedural fluency, and strategic competence in mathematics while modeling the pedagogy touted to be effective at leading to mathematical understanding in students.

Middle grades teachers comprised a large number of participants in the *Patterns* course along with some elementary, secondary, and pre-service teachers. A number of IHE faculty members were also full participants in the course alongside the in-service teachers. Response to the course was highly positive and, subsequently, IHE faculty members made revisions at two IHEs to include the content and pedagogy of *Patterns* in a course for pre-service teachers. The evaluation of the *Patterns* summer course was focused on the middle grades teachers.

When evaluating the effectiveness of professional development in mathematics education, one of the most challenging aspects is finding or developing instruments to measure changes in teachers' content knowledge. Typically, content knowledge in mathematics is viewed rather narrowly as an ability to arrive at an accurate answer to a mathematical problem. Very rarely is there an interest in examining the problem-solving process, including misconceptions and verification, as a pathway to understanding. This evaluation takes a more expansive view of mathematical content knowledge and uses a variety of data sources that provide a more complete picture of teachers as learners and teachers of mathematics than what could be gleaned from only a multiple choice test. Teachers' work was examined not only for accuracy of final answers, but also for processes in obtaining answers and a subsequent verification process involving articulating sound reasoning for the answers obtained.

2. Methodology

***Objective Assessment of Content Knowledge*.** The Learning Mathematics for Teaching (LMT) project at the University of Michigan has developed and made available several sets of items designed to assess content knowledge for teaching mathematics. Currently, items are available for elementary mathematics and middle school mathematics in the areas of (a) patterns, functions, and algebra; (b) geometry, and (c) numerical reasoning. The evaluation staff worked with Mathematics Education

Collaborative (MEC) instructors to select items from the item pools that best matched the *Patterns, Functions, and Algebraic Reasoning* course taught in Summer 2005. A few items were considered a very good match with course content, some items matched moderately well, and other items did not match at all. It was decided to use only items that fell into the first two categories and exclude those items that did not match. Although the match in some cases was only moderate, it was felt that the resulting scale should show improved content knowledge as a result of the course.

Using the item parameters provided by LMT staff, the resulting test (comprised of 36 items) was analyzed. Although the LMT items are grouped into scales according to the areas listed above, each of which has been extensively analyzed, the LMT project permits use of selected items to build scales better suited to particular circumstances. This scale, the Content Knowledge for Teaching Mathematics – Patterns (CKTM-Patterns) was used an objective measure of increased content knowledge that occurred as a result of participation in a Summer 2005 (Year 1) course. The CKTM-Patterns was administered to middle school teachers and pre-service teachers as a pre-test at an orientation session prior to each course session. The same instrument was administered again on the last day of the course to the same individuals. The pre-tests and post-tests for individuals were then paired so that individual progress could be examined. For further evidence of reliability, a Cronbach alpha was calculated to determine the internal consistency of the 36-item test. The test showed good internal consistency (alpha = 0.82).

Following the Year 1 administration, the CKTM was modified. The evaluation team met with the course instructors and project management team to review the items again for content validity. It was hypothesized by the team that the test was taking participants too long to complete. As well, some items were viewed by the instructors as unrelated to course content. As a result, seven items were removed from the test. The test maintained its integrity, and the new set of 29 items still showed good internal consistency (alpha = 0.80)

After further deliberation with the project team and instructors, two open-ended items were selected from a group of items developed by Nanette Seago, Co-PI, Turing to the Evidence and Video Cases II Projects, WestEd. These two items required participants to generate an equation to describe a situation. The items were determined by the team and instructors to be an excellent fit with course content and were added to the CKTM. The Year 2 CKTM-Patterns contained 31 items.

Performance Assessment. During the summer courses, instructors administered a performance assessment at the beginning of the course and again at the end of the course. During Year 1, for courses in the first session, two different assessments were administered and could not, therefore, be compared. During the second and third session courses of Year 1, however, the same assessment was administered pre and post. Evaluators used the Oregon Department of Education Mathematics Problem Solving Official Scoring Guide as their rubric to assess four domains (conceptual understanding, processes & strategies, communication, and accuracy). Only those assessments for which both the pre-task and the post-task were available were scored.

Because of the inability to use any assessments from the first session and the lack of matching pre and post assessments, only 16 performance assessments were scored in Year 1. However, 70 performance assessments were scored in Year 2.

Although the rubric has been used successfully in Oregon, the evaluators took precautions to ensure that there was consensus on the performance assessment scores. Three scorers applied the rubric individually to each of the assessments then met to discuss their scores and to resolve any discrepancies. The individual results were in good agreement in most cases, and in all cases, the consensus scores were easily determined.

Course Portfolio. As part of the 2005 summer courses, each participant kept a portfolio of completed tasks, assignments, and reflections. The portfolios were assessed in the spring of 2006 using a rubric designed by the evaluation team that was based on the project definition of "challenging courses and curriculum". Three members of the evaluation team piloted the rubric on five portfolios and, in January 2006, the rubric was reviewed by the rest of the GBMP team. Slight modifications were made as a result of both the pilot exercise and input from the team.

Five portfolios from middle school teachers were randomly selected from each of the three summer course sessions, resulting in a total of fifteen portfolios to be scored. The same process of reaching consensus on scores that was applied to the performance assessments was used with the portfolios. The same procedure for sampling and scoring portfolios was applied to a revised university mathematics course.

Behavioral Checklist. The CEA developed a behavioral checklist for use in observing teachers as they participate in the GBMP courses. The checklist contains four dimensions—deepening mathematics understanding, productive disposition, inquiry and reflection, and communication. These four dimensions are the same dimensions that form GBMP's definition of challenging courses and curriculum. For each dimension, there are four observable behaviors that were determined to be evidence of each dimension by the evaluation team. The checklist was piloted during the Summer 2006 course. Three course participants were selected randomly to be observed on three occasions throughout the two-week course.

Course Satisfaction Survey. The last day of course sessions was used to collect data from the participants. One of the instruments used was a simple course satisfaction survey that was comprised of seven Likert-type items and four open-ended items. The items address participants' perceptions of how the course may have benefited them and how well the course was presented. All participants were asked to complete this instrument.

Classroom Observations. The Reformed Teaching Observation Protocol (RTOP), specifically designed for mathematics and science class observations, was used as the classroom observation tool. The instrument focuses on five areas: (1) lesson design and implementation, (2) propositional knowledge, (3) procedural knowledge, (4) communicative interactions, and (5) student/teacher relationships.

In addition to the middle school classrooms, the RTOP was also used in five undergraduate mathematics classrooms. Four of the university mathematics courses were traditional introductory university mathematics courses; however, one had been redesigned to incorporate the strategies demonstrated in the *Patterns* course.

3. Results

Content Knowledge for Teaching Mathematics-Patterns**.** The CKTM-Patterns was used to monitor the content learning gains of teachers during the 2005 and 2006 summer classes. In Year 1, pre-test and post-test paired results were available for 62 practicing teachers and four pre-service teachers, eleven of whom were recruited to serve as Math Support Team (MST) teachers after the summer course (MSTs are teachers who will gradually take on leadership roles within their schools and districts).

The baseline performance of MSTs was higher than that of other teachers and of the pre-service teachers. On average, pre-service teachers and MSTs passed one additional item on post-test than they passed on pre-test. Although other participants started lower than MSTs, on average, they gained approximately two points between pre-test and post-test.

In Year 1, raw scores for $n = 66$ teachers who had both pre-test and post-test scores were examined using the effect size index. The teachers' pre-test scores served as their own controls. The difference between the post-test score arithmetic mean and the pre-test score arithmetic mean formed the numerator, and the standard deviate of the pre-test score was used as the denominator. These statistics were: xbar_pre = 21.08, xbar_post = 23.20, and sigma_pre = 5.82.

The differences in means between post-test and pre-test yielded an effect size of 0.34, which falls midway between the effect size index for "small" effects, 0.2, and the index for "medium" effects, 0.5, (Cohen, 1988). Educational research often considers effect sizes in excess of 0.33 to be practically meaningful.

Table I: *Year 1 CKTM Pre-Post Results (36 items)*

GROUP	Pre-test	Post-test
All Participants *(n = 66)*	21.08	23.2

For additional interpretation, three different **U** statistics were tabled. These were based on the probabilities derived by the area under the standard normal curve. The $\mathbf{U_3}$ statistic is equal to the cumulative probability under the normal curve from negative infinity to the calculated effect size index. In Year 1, when the effect size is 0.34, the cumulative probability is 0.6331. In other words, the upper-half of the post-test score population exceeds 63.31% of the pre-test score population.

Table II: *Year 2 CKTM Pre-Post Results (31 items)*

GROUP	Pre-test	Post-test
All Participants *(n = 98)*	16.04	18.78

In Year 2, raw scores for $n = 98$ teachers who had both pre-test and post-test scores were examined. The difference between the pre-test score arithmetic mean (16.04) and the post-test score arithmetic mean (18.78) yielded an effect size of .46. This index falls at the high end of the effect size index for small effects, 0.2, and close to the index for medium effects, 0.5, (Cohen, 1988). When the effect size is 0.46, the cumulative probability is 0.6772, meaning the upper-half of the post-test score population exceeds 67.72% of the pre-test score population.

Because the last two items on the CKTM were such a good fit with course content and because they were open-ended items, they were analyzed separately. For $n = 98$ teachers, on the first of the two items, there was over a 30% increase in the number of teachers who responded correctly from the pre-test to the post-test. On the second item, there was close to a 40% increase in the number of teachers who responded correctly.

Table III: *CKTM Pre-Post Scores on Open Ended Items*

CKTM Item Number	% Teachers with Correct Response on Pre-Test	% Teachers with Correct Response on Post-Test
30	29.7%	63.4%
31	33.7%	73.3%

One additional change noted in participants' responses to the last two CKTM items in Year 2 was an increase in the amount of work shown by the participants from pre-test to post-test. On the pre-test, some participants wrote nothing, others wrote only an equation with no explanation, and others scribbled a few notes as they tried to solve the problems. On the post test, most participants recorded their problem-solving process in enough detail so that an outside evaluator could follow their thinking from translating the problem from pictures to numbers to generating an equation to capture the pattern presented. It is believed by the partnership that this represents powerful evidence of change, and the evaluation team is currently in the process of developing a rubric to capture that change more systematically.

Performance Assessment. In Year 1, sixteen teachers were scored on a pre-post performance assessment task similar to the tasks used as part of the 2005 summer courses. These tasks were evaluated using the Oregon scoring rubric for performance assessments in mathematics.

Results revealed that, even at pre-test, accuracy was quite high, but accuracy still improved by post-test. Significant and impressive growth occurred in all other areas. The management team and evaluation team discussed whether these results were a function of "teaching to the rubric" and identified potential mechanisms for minimizing this threat to validity in subsequent summer course assessments.

Table IV: *Year 1 Performance Assessment Pre-Post Score Frequencies*

Score	1		2		3		4		5	
	N Pre	N Post	N Pre	N Post	N Pre	N Post	N Pre	N Post	N Pre	N Post
Processes and Strategies (PS)	5	0	6	0	3	1	2	7	0	8
Conceptual Understanding (CU)	3	0	8	0	2	2	2	5	1	9
Communication (C)	4	0	6	0	4	2	1	6	1	8
Accuracy (ACC)	4	1	N/A	N/A	N/A	N/A	10	2	2	13

Table V: *Year 2 Performance Assessment Pre-Post Score Frequencies*

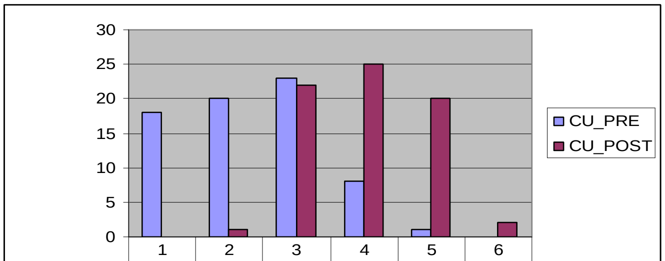
Score	1 – 1.5		2 – 2.5		3 – 3.5		4 – 4.5		5 – 5.5		6	
	N Pre	N Post	N Pre	N Post	N Pre	N Post	N Pre	N Post	N Pre	N Post	N Pre	N Post
Processes and Strategies (PS)	20	0	24	4	21	19	4	28	1	19	0	0
Conceptual Understanding (CU)	18	0	20	1	23	22	8	25	1	20	0	2
Communication (C)	22	0	28	3	14	19	3	28	3	16	0	4
Accuracy (ACC)	21	5	N/A	N/A	N/A	N/A	39	26	10	39	N/A	N/A

Table VI: *Year 2 Performance Assessment Pre-Post Median Scores*

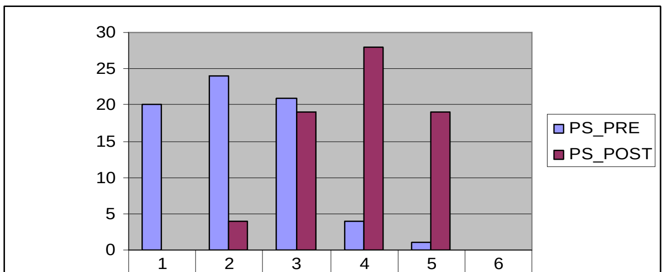
N=48	Conceptual Understanding (CU)		Processes and Strategies (PS)		Communication (C)		Accuracy (ACC)	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Median Score	2.000	4.000	2.000	4.000	2.000	4.000	4.000	5.000

The median score increased from a 2.0 to a 4.0 on three of four dimensions. The descriptors for performance at the 2.0 level included underdeveloped, sketchy, ineffective, and unclear. Descriptors for performance at the 4.0 level included complete, adequate, relevant, explained, and supporting the solution.

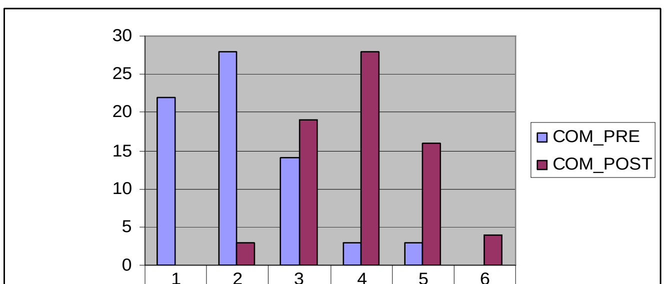
Figure I: *Year 2 Performance Assessment*
Pre-Post Score Frequencies by Rubric Dimension



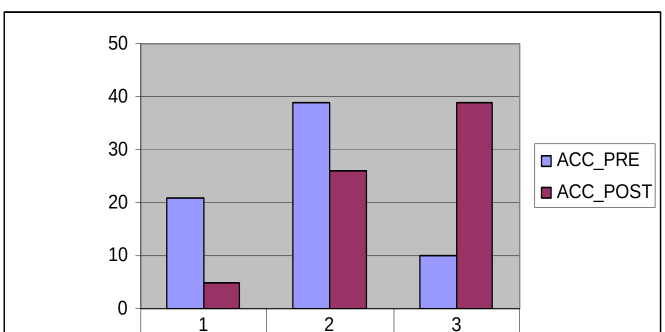
	1	2	3	4	5	6
CU_PRE	18	20	23	8	1	0
CU_POST	0	1	22	25	20	2



	1	2	3	4	5	6
PS_PRE	20	24	21	4	1	0
PS_POST	0	4	19	28	19	0



	1	2	3	4	5	6
COM_PRE	22	28	14	3	3	0
COM_POST	0	3	19	28	16	4



	1	2	3
ACC_PRE	21	39	10
ACC_POST	5	26	39

A Wilcoxon Signed Ranks Test showed significant improvement on all four dimensions from pre to post administration. Participants showed marked improvement in the areas of conceptual understanding, processes and strategies, and communication, and some improvement on accuracy, where there was little room for improvement because of the limited range of rubric scores for that dimension.

Table VII: *Year 2 Wilcoxon Signed Ranks Test*

	POST Conceptual Understanding - PRE Conceptual Understanding	POST Processes & Strategies – PRE Processes & Strategies	POST Communication - PRE Communication	POST Accuracy - PRE Accuracy
Z	-6.629(a)	-6.779(a)	-6.926(a)	-4.983(a)
Asymp. Sig. (2-tailed)	.000(b)	.000(b)	.000(b)	.000(b)

a Based on negative ranks.
b Wilcoxon Signed Ranks Test

Portfolios. Other evidence of teacher understanding can be derived from an analysis of portfolios generated during summer courses. Teachers participating in the 2005 summer course completed portfolios as part of their participation. Portfolios contained the following components: reflective pieces on the teacher as a learner of mathematics and a teacher of mathematics; letter to someone addressing the “big ideas” from the course and changes anticipated to be made in the classroom; pre-assessment task; self-selected “most important piece of work” from the course; scored task from the course; teacher-selected task believed to reflect the teacher as a learner of mathematics; and an assigned assessment task from the course.

A sample of portfolios was rated holistically in terms of five features, all related to the challenging courses and curriculum (CCC) dimensions delineated previously by the design team. The first two dimensions relate to the CCC feature of deepening knowledge of big ideas in mathematics. The other dimensions directly relate to the remaining three components of challenging courses and curriculum. Results of consensus judgments among three raters are provided in the table below. These ratings indicate that more than half of the sample demonstrated performance that was at or above the proficient level on each dimension. Year 2 portfolios have not yet been scored.

Table VIII: *Year 1 Summer Course Portfolio Scores*

	Median	N Incomplete	N Emerging	N Proficient	N Expert
Problem Translation	3.00	0	4	6	5
Mathematical Procedures	3.00	0	2	9	4
Productive Disposition	3.00	1	1	9	4
Inquiry and Reflection	3.00	1	3	7	4
Justification and Communication	3.00	1	5	5	4

Similar results were found for the University of Alabama at Birmingham revised mathematics course. Median portfolio performance was at the proficient level for pre-service teachers. More impressive was that none of the pre-service teacher portfolios were scored at the incomplete level.

Table IX: *Revised UAB Course Portfolio Scores*

	Median	N Incomplete	N Emerging	N Proficient	N Expert
Problem Translation	3.00	0	1	6	3
Mathematical Procedures	3.00	0	1	7	2
Productive Disposition	3.00	0	1	5	1
Inquiry and Reflection	3.00	0	2	6	2
Justification and Communication	3.00	0	2	5	3

Behavioral Checklist. The checklist was piloted in Year 2, in the Summer 2006 *Patterns* course. Three course participants were chosen at random to be observed three times over the two-week course—once on the first day (T1), once on the fourth day (T2), and once on the eighth day (T3). Observations took place when participants were working in groups or working on menu tasks with other participants. For each of the three participants, change occurred on all four dimensions. In the table below, an "X" is used to indicate when a behavior is observed.

Table X : *Behavioral Checklist Results for Selected Participants by CCC Dimension*

	Participant 1			Participant 2			Participant 3		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
Understanding of Mathematical Ideas									
Uses variables to describe unknowns		X	X			X	X	X	X
Explains why equations make sense geometrically			X			X		X	X
Represents linear and quadratic equations in variety of ways			X			X		X	X
Productive Disposition									
Persists when answer is not known		X	X			X		X	X
Asks for guidance but not answers			X			X		X	X
Tries variety of strategies for approaching problem			X			X			X
Inquiry and Reflection									
Makes extensions and connections beyond immediate problem			X					X	X
Explores why it works and whether it will always work									X
Confusion and mistakes lead to further exploration			X			X			X
Communication									
Explains reasoning fluently			X			X			X
Asks probing questions			X						X
Shares ideas with class		X	X			X		X	X

Course Satisfaction Survey. Of the 143 participants who took the *Patterns* course and completed the survey, 98.6% agreed or strongly agreed that the course improved their mathematical skills and understanding. The same percentage of participants agreed or strongly agreed that the mathematical ideas presented in this course would be useful in their teaching. And 100% of the participants agreed or strongly agreed that the course improved their understanding of pedagogy.

Participants were asked to respond to open-ended questions about their learning and growth during the *Patterns* course. Teachers' quotes from those open-ended items follow:

> "The most important thing that I am taking with me from this course is how to effectively teach mathematics. I have always known that this would be a better way to guide students' understanding but wasn't sure how and didn't have the resources. I feel like I'm now equipped with both! Also, I now understand algebra! If I had learned it this way to begin with, life would have been a lot simpler! This course has been a true asset to me personally and professionally!"

> "I cannot believe how much stretching I was forced to do in this course. The course has improved my math reasoning skills. Now I actually have fun figuring out these problems - because I now know I can! My growth in understanding math has been enormous!"

> "It was important for me to see how students see mathematics in different ways. I have always shown my students more than one way to work problems but I did not truly understand how people can have the same situations and come up with the same answer, but look at the situations in so many different ways. I definitely see a big difference in my understanding of mathematics because I was able to make connections with principles that I knew but actually construct them and make sense of why the numbers worked."

> "This course really inspired me. I felt as many of my students do when approached with a mathematical thinking problem. I felt unsure and scared. I didn't really feel like I had a strong grasp of mathematical thinking when it came to Algebra. I now notice patterns and make connections to other things. I feel I can understand. I no longer have to just accept an answer or strategy because the instructor tells me that's how it is done. I can take ownership of my own learning and come to an understanding for myself."

Classroom Observations. A sample of middle school mathematics classrooms has been observed each school year. Attempts were made to capture repeated observations of the same teachers. Although some repeated observations were conducted, logistical problems (teacher attrition, scheduling, teacher participation) largely prevented teacher follow-up. Therefore, presented below are the results of separate samples of classroom teachers. Year 1 observations (n = 11) were conducted in classrooms with teachers prior to their completion of the *Patterns* courses. Observations from Year 2 (n = 12) were conducted with teachers who had all taken a *Patterns* course. Year 3 observations (n = 40) include some teachers who had taken only *Patterns* and some teachers who had taken more than one summer course. Two observers visited each class.

Table XI: *Reformed Teaching Observation Protocol in District Classrooms*

RTOP Categories (maximum score on each subscale is 20)	Year	Median
Lesson Design/Implementation	Year 1 Year 2 Year 3	4.00 11.75 14.50
Propositional Knowledge	Year 1 Year 2 Year 3	5.50 10.75 13.25
Procedural Knowledge	Year 1 Year 2 Year 3	4.00 13.00 14.00
Communicative Interaction	Year 1 Year 2 Year 3	4.00 12.75 14.00
Student/Teacher Relationships	Year 1 Year 2 Year 3	5.50 14.00 15.50

These results from Year 1 suggest baseline classroom contexts that were largely teacher-directed, didactic, whole-group, and focused on delivery rather than inquiry and discourse. The results from Years 2 and 3 show substantial improvement on each of the scales.

Observations were also conducted in IHE faculty classrooms. Four traditional undergraduate mathematics courses were observed. RTOP scores from those courses were compared to RTOP scores from an observation of the UAB course that was revised to include the content and pedagogy of the *Patterns* course. Results are presented below.

Table XII: *Reformed Teaching Observation Protocol in UAB Courses*

RTOP Categories (maximum score on each subscale is 20)	Traditional UAB Courses Median	(Range)	Revised UAB Course
Lesson Design/Implementation	1.0	(0 - 3)	18
Propositional Knowledge	3.0	(3 - 12)	11
Procedural Knowledge	2.0	(0 - 6)	16
Communicative Interaction	1.0	(0 - 3)	14
Student/Teacher Relationships	2.0	(0 - 7)	15

As with the middle school teachers' classrooms, results from IHE faculty mathematics courses suggest classroom contexts that were largely teacher-directed, didactic, whole-group, and focused on delivery rather than inquiry and discourse. In contrast, the mathematics education course revised to reflect the summer *Patterns* course yielded RTOP scores consistent with the standards of inquiry-based, reformed mathematics instruction.

In conclusion, participants in the inquiry-based mathematics courses demonstrated gains in all four CCC dimensions: (1) deepening knowledge of important mathematical ideas; (2) productive disposition; (3) inquiry and reflection; and (4) communication. This growth was evidenced by a variety of data sources. Self-reported data echoed these results as participants described increased understanding of mathematics and increased confidence as users and teachers of mathematics. Participants characterized the inquiry-based course as a powerful learning experience and classroom observations indicated a shift in classroom practice toward providing a more inquiry-oriented experience for their own students.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd. ed.), Hillsdale, NJ: Lawrence Erlbaum.

Contact Information

Rachel Cochran, GBMP Lead Evaluator, danelle@uab.edu

John Mayer, GBMP principal investigator, mayer@math.uab.edu

Bernadette Mullins, GBMP co-investigator, bmullins@bsc.edu

The Greater Birmingham Mathematics Partnership is funded by the National Science Foundation through Cooperative Agreement #0632522. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.