**Knowledge, Pedagogy, Practice or Student Achievement:**

**Evaluating a Middle School Math M.Ed. Professional Development Program.**

**David Ensminger**

**Jim Breunlin**

**Lindsay Vielleiu**

**PROGRAM DESCRIPTION**

**Program Origins**

Loyola’s Middle School Math M.Ed. program was funded through a grant received from the Illinois Board of Education. The granting agency provided two phases of funding, initial funding included monies for the design and development of a professional development program aimed at increasing knowledge and skills of teachers in the areas of math content and math pedagogy.

Universities that were successful in developing program proposals received funding for program implementation. The granting agency established requirements for program proposals. These include partnerships with Local Educational Associations (LEAs), partnerships within the University and strict guidelines for program evaluation plans.  These standards required that evaluation plans adhere to GPRA guidelines, that either experimental or strong quasi-experimental designs be employed, utilize recommended instruments for content knowledge assessment, Diagnostic Teacher Assessments in Mathematics and Science (DTAMS), and that student achievement gains be assessed using the standardized student assessment, Illinois Standards Achievement Test (ISAT). These measures were the prescribed means for judging program effectiveness. Less emphasis was placed on evaluating the outcomes such as changes in pedagogical perspectives or pedagogical practices. Informal communications from the funding agency indicated that only proposals with strong evaluation plans that meet the agencies guidelines would be funded.

Data obtained from the LEA indicated that many teachers who taught Math in middle schools for the LEA did not possess a math teaching endorsement.  Based on this observation, Loyola University Chicago submitted a program proposal that expressed the views of the program designers - that a program which led to both a Masters of Education and a math teaching endorsement would provide participants with the necessary knowledge and skills in both content and pedagogy to impact student learning and achievement in Mathematics.

This proposal defined a collaborative relationship between the LEA, Loyola’s School of Education, (LSOE) and Loyola’s College of Arts and Sciences (LCAS). The program was designed to enroll two cohorts of 15-20 participants simultaneously with LSOE paying two thirds of the participants’ tuition and LEA paying one third.

Based on this program design the initial evaluation called for the use of a cohort evaluation model that used each cohort group as control group for the other in order to evaluate the content knowledge outcomes and pedagogical perspectives.  Content knowledge was to be assessed using the DTAMS and pedagogical perspectives measured using a modified version of the Classroom Mathematics Teachers Questionnaire, developed by the Horizon Research Inc. Outcomes related to student achievement were to be determined using a non-equivalent control group design that utilized local teachers who were working in the same school as the program participants to serve as the control group. The evaluation plan included an observation and interview component that would allow for data collection on the actual changes participants were making in their practice along with a control group to compare observations and interview data.

**Program Changes and Evaluation Plan Modifications**

The program enrolled only 25 teachers currently working in schools of the LEA. Given the limited number of program participants, the LSOE and LCSA determined that they would not be financially able to staff two separate cohorts; therefore, the proposed cohort evaluation design was modified. The design was modified to a single group pretest posttest design to measure content knowledge outcomes, and a non-equivalent control group design with repeated measures for both the pedagogical practice and the student achievement outcomes. The observation and interviews remained as originally proposed.

Within the first year of program implementation the LEA withdrew their financial support for the program (Note: funding from LEA was withdrawn for all local MSP partnerships). This reduction in funds brought up major concerns for the program and resulted in proposed changes to the evaluation plan, particularly the loss of a graduate assistant to support evaluation activities. This specifically resulted in questions being raised concerning the collection of observation and interview data in order to evaluate the program’s outcomes and value. In addition to the loss of LEA financial support, the program director for the project stepped down from her role leaving the coordination of the project activities and support of the evaluation activities related to the collection of content knowledge and student achievement data collection in jeopardy.

The newly appointed program director and the lead evaluator met to determine the best course of action given the changes in the program’s budget related to personnel and evaluation work.  While the funding agency expressed the most concern with the continued collection of student achievement data, the lead evaluator and the program director determined that the proximal outcomes related to content knowledge, pedagogical perceptions and pedagogical practice were more valuable to determine the program’s successes and program improvement.  In order to allow for data collections on proximal outcomes the program director and lead evaluator determined that a graduate assistant was needed to support evaluation activities and specifically collect observation and interview data related to teacher’s practices. The program director modified subsequent budgets in order to hire a graduate assistant for 2 years.  The final changes to the evaluation plan involved the removal of the control group and the switching to single groups pretest posttest design for pedagogical perspectives.  This collaboration between the new program director and the lead evaluator resulted in a more open dialogue for evaluation purposes and formation of an evaluation team consisting of the lead evaluator, the program director and the graduate assistant.

**PROGRAM OUTCOMES**

*Student Achievement*:
 Illinois Standards Achievement Test (ISAT) data was compared for 2010 and 2011.  Candidates who taught mathematics both years were included in the analysis.  Student data collected in candidate classrooms are mixed in results.  Overall there was no significant difference between the 2010 mean ISAT Math score of 248.7 when compared with the 2011 score of 249.1 (t= .391, p = .377 ).   One must remember that there was a significant difference between the 2009 and 2010 scores overall and at all grade levels and the 2011 scores have remained steady compared to 2010.      At the 6th grade level, there was a statistically significant difference between the  2010 ISAT Math mean of 236 and the 2011 mean of 251 (t=5.06, p=.000).   There was no statistically significant difference at the 5th grade level ( 2010 mean 252, 2011 mean 249, t=.75, p=. 231) nor the  8th grade levels:  ( 2010 mean 246, 2011 mean 245, t = .408, p=.341).  At the 7th grade level the 2010 mean of 264 was significantly greater than the 2011 mean of 258 ( t = 2.79, p =. 003)

*Content Knowledge*:

During the program candidates were enrolled in several math content courses: Probability and Statistics, Geometry, Algebra and Number Theory. Each of these courses infused science applications related to each of the core mathematical areas. Content knowledge outcomes were evaluated using the DTAMS in a pre/ posttest analysis. The DTAMS provides a composite knowledge score and four individual knowledge scores. Each individual knowledge score relates to a specific type of learning. Knowledge Type scores: Type I – Memorized/Factual Knowledge; Type II – Conceptual Understanding; Type III – Reasoning/Problem Solving; Type IV – Pedagogical. Table 1 provides the results for the pre/post analysis for the overall score, and the four knowledge types for each content area.

The analysis of DTAMS scores for Algebraic idea, Number Computing and Probability and Statistics, indicate that candidates showed significant gains in their composite knowledge posttest scores, indicating that the program courses associated with these content areas increased candidates’ knowledge. Additionally, Type I-III individual knowledge scores were significantly different for these three DTAMS analyses. Indicating that for each of these three courses, candidates showed significant improvement in their factual, conceptual and problems solving knowledge related to these content areas. The Type IV pedagogical knowledge posttest scores were statistically significant for Number Computation and Probability and Statistics, indicating that these courses increased candidates’ knowledge related to the teaching of this content. The pedagogical knowledge score for Algebraic Ideas approached significant *p* = .053, suggesting that the program had some impact on candidates’ knowledge surrounding the teaching of this content.

The analysis of the DTAMS post scores for Geometry and Measurement indicated no significant change for the overall knowledge and individual knowledge scores. While on the surface this would indicate that the Geometry course had no impact on the candidates’ knowledge, the instructor of the Geometry course indicated that the course focused solely on Geometry and the DTAMS included questions specifically designed to assess measurement facts, concepts, and problem solving and teaching. This suggests that the DTAMS may not have been a valid measurement for this course. These results provide evidence of candidates’ mastery of the content and that the program successfully achieved the proximal outcome of increasing the candidates’ content knowledge in mathematics.

Table 1. *Total and Individual Knowledge Sores for DTAMS*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean | SD | N | t | p |
| Algebraic Ideas | Pre-Total Knowledge Score | 16.07 | 7.37 | 15 | -6.248 | .000\* |
| Post Total Knowledge Score | 24.47 | 6.98 | 15 |
| Pre-Knowledge Type I | 4.27 | 2.69 | 15 | -5.054 | .000\* |
| Post-Knowledge Type I | 7.27 | 2.15 | 15 |
| Pre-Knowledge Type II | 4.93 | 2.49 | 15 | -3.944 | .001\* |
| Post-Knowledge Type II | 6.93 | 2.12 | 15 |
| Pre-Knowledge Type III | 4.07 | 1.49 | 15 | -5.533 | .000\* |
| Post-Knowledge Type III | 6.53 | 2.17 | 15 |
| Pre-Knowledge Type IV | 2.80 | 2.08 | 15 | -2.114 | .053 |
| Post-Knowledge Type IV | 3.73 | 2.15 | 15 |
|  | Mean | SD | N | t | p |
| Geometry/Measurement | Pre-Total Knowledge Score | 16.13 | 6.185 | 16 | -.780 | .447 |
| Post Total Knowledge Score | 17.19 | 7.626 | 16 |
| Pre-Knowledge Type I | 5.19 | 1.834 | 16 | -1.576 | .136 |
| Post-Knowledge Type I | 5.81 | 1.834 | 16 |
| Pre-Knowledge Type II | 5.31 | 1.352 | 16 | 1.732 | .104 |
| Post-Knowledge Type II | 4.56 | 1.931 | 16 |
| Pre-Knowledge Type III | 2.56 | 2.128 | 16 | .126 | .901 |
| Post-Knowledge Type III | 2.50 | 2.129 | 16 |
| Pre-Knowledge Type IV | 3.06 | 2.516 | 16 | -1.966 | .068 |
| Post-Knowledge Type IV | 4.31 | 2.750 | 16 |
|  | Mean | SD | N | t | p |
| Number Computation | Pre-Total Knowledge Score | 15.17 | 5.894 | 18 | -6.665 | .000\* |
| Post Total Knowledge Score | 27.17 | 6.214 | 18 |
| Pre-Knowledge Type I | 5.17 | 1.724 | 18 | -6.775 | .000\* |
| Post-Knowledge Type I | 8.17 | 1.581 | 18 |
| Pre-Knowledge Type II | 4.94 | 1.474 | 18 | -3.465 | .003\* |
| Post-Knowledge Type II | 6.67 | 1.572 | 18 |
| Pre-Knowledge Type III | 2.44 | 2.007 | 18 | -6.039 | .000\* |
| Post-Knowledge Type III | 5.67 | 2.086 | 18 |
| Pre-Knowledge Type IV | 2.61 | 2.304 | 18 | -5.333 | .000\* |
| Post-Knowledge Type IV | 4.31 | 2.750 | 16 |
|  | Mean | SD | N | t | p |
| Probability and Statics | Pre-Total Knowledge Score | 15.63 | 5.277 | 16 | -3.943 | .001\* |
| Post Total Knowledge Score | 19.50 | 7.616 | 16 |
| Pre-Knowledge Type I | 5.94 | 1.611 | 16 | -3.569 | .003\* |
| Post-Knowledge Type I | 6.75 | 1.483 | 16 |
| Pre-Knowledge Type II | 4.81 | 1.721 | 16 | -1.568 | .138 |
| Post-Knowledge Type II | 5.63 | 2.062 | 16 |
| Pre-Knowledge Type III | 3.31 | 2.798 | 16 | -2.267 | .039\* |
| Post-Knowledge Type III | 4.19 | 2.401 | 16 |
| Pre-Knowledge Type IV | 1.50 | 1.317 | 16 | -2.206 | .043\* |
| Post-Knowledge Type IV | 2.94 | 2.645 | 16 |

\*p < .05

 As part of the data collection process, a graduate student member of the evaluation team conducted classroom observations of MAST candidate students currently enrolled in the Middle School Math program and teaching mathematics in their schools. Candidates were observed teaching a math class; the class levels ranged from the fifth to eighth grade. The observations took place in spring of year one and year two of the program. The data collection instrument used was Inside the Classroom Observation and Analytic Protocol. Following the observation a brief teacher interview was conducted.  Since the instrument uses a Likert scale, the nonparametric  Mann-Whitney U test was used to detect differences between year one and year two.  No significant differences were found, but this may be due to the small sample size that yields low Power and a higher probability of Type 2 error.

However, teacher interviews post-observation indicated a high confidence with mathematical terminology, procedure and application of mathematical concepts among the MAST candidates.  One hundred percent of the candidates reported that their recent M.Ed. coursework had improved their knowledge and confidence related to mathematics.

A focus group was conducted with the MAST candidates in May 2011.  Many candidates reported a sharp increase in their mathematical confidence and efficacy related to their understanding of mathematics. Participants made the following statements indicating an increased level of efficacy related to their math content knowledge: “[I have ] less reliance on textbooks and computations.  I can play math games and discuss why they work”; “I am confident with the math material which makes me more comfortable answering student questions”; “ I can guide students to discover math for themselves … Even if I can not immediately answer a question, I am confident enough now to say let’s figure it out together … and we do”; “I can integrate science and math; make that connection in science class with math”. The qualitative data from interviews and focus groups substantiates the DTAMS results that the program achieved its desired outcome of increasing the content knowledge. Although not directly measured, comments from candidates during interviews and the focus group indicate that the program also impacted their self-efficacy related to math, (i.e. their personal belief that they can do math). While not a direct measure of content knowledge, increased levels of self efficacy related to a subject is associated with increased risk taking and willingness to engage in actions related to the subject. This increased efficacy suggests that the candidates may continue their own learning related to math and continue to increase their content knowledge post program.

Results from the DTAMS, observations, interviews and focus group provide significant evidence that the Middle School Math program impacted the knowledge content of the candidates. These results also provide evidence of the program’s success related to the professional development of math teachers’ as content knowledge of teachers is associated with student achievement.

*Pedagogical Perspectives*:
 The evaluation team also considered a change in the ways candidates viewed the teaching of math as a critical outcome for the program and an outcome needed in order to facilitate the shift in actual pedagogical practice in the classroom. Changes in pedagogical perspectives were evaluated though an online pedagogy perspectives questionnaire at the beginning of the program (May 2008) and at the end of the program (May 2011).  A pairwise analysis was done on this nonparametric ordinal data using a Mann-Whitney U Test.  There were several areas that revealed a statistically significant increase.  Table 2. shows the areas where pedagogical perspective shifted during the course of the M.Ed. Program. There were nine other responses where a statistically significant increase existed.  This provides considerable evidence that the candidates’ teaching practice changed during the two-year program.

Table 2. *Changes in Pedagogical Perspectives Measured by Pedagogical Questionnaire*

|  |  |
| --- | --- |
| Pedagogical Perspectives | P value |
| Take students' prior understanding into account when planning curriculum and instruction | .022 |
| Have students work in cooperative groups | .033 |
| Make connections between mathematics and disciplines other than science | .027 |
| Develop students’ conceptual understanding of mathematics | .028 |
| Use of technology in support of mathematics instruction | .041 |
| Target problem solving | .026 |
| Target reasoning and proof | .018 |
| Use of Multiple Representations (e.g. concrete models, and numeric, graphical, symbolic, and geometric representation | .002 |
| Leads a class using investigative strategies | .035 |
| Recognize and respond to student diversity | .008 |
| Encourage student interest in mathematics | .029 |
| Ask students to consider alternative methods of solution | .008 |

 Observations and interviews also evidenced outcomes related to changes in pedagogical perspectives. One candidate stated the Middle School Math program resulted in the candidate asking her students more questions and encouraging students to actively think and question the math material independent of her inquiries. Another candidate explained the program taught her to begin math instruction using concrete objects students could manipulate and then move toward more abstract methods in which students only use numbers. Another stated the program aided her in becoming more creative in her teaching methods, encouraging her to design a project in which her students would assume ownership and enjoy resulting in the development of a data collection group project that was observed by the evaluator. One candidate stated that even when she uses a lesson directly from a curriculum, the program has affected her approach to the lesson; she has become more comfortable making adjustments in the middle of a class activity if the students are not understanding the material. Another candidate spoke of how the program increased her confidence to teach lesson plans that encourage the students to actively participate in non-conventional ways. Two candidates spoke of how the program caused an awareness of the importance of inquiry-based teaching; “… it is not just important for the student to answer a problem correctly, it is equally important to understand the underlying logic and how they arrived at the answer.”

Changes in pedagogical perspectives were also evident in comments from participants during the focus group, “I can connect math to other subjects (history; history of famous mathematicians); makes math more exciting” “Math is not an isolated subject; we now look at it as an integrated part of multiple subjects; that there is a need to make math relevant/real-world applied in order to make interesting and meaningful to students” “I can integrate science and math; make that connection in science class with math.” These statements indicate that participants’ now view the teaching of math from an integrative perspective and that math as a subject area has relevance and application beyond the math class.

Results from the pedagogical questionnaire analysis, observations, interviews and focus group provide significant evidence that the Middle School Math program impacted the pedagogical perspectives of the candidates. These results also provide evidence of the program’s success related to the professional development of math teachers, as increased pedagogical knowledge is often associated with student achievement.

*Pedagogical Practices*:

Classroom observation revealed students in MAST candidates’ classrooms are likely to be engaged in doing mathematics, use cooperative groups, and participate in activities that use inductive inquiry or guided discovery.  These practices target conceptual understanding.   More specifically, 77 % of candidates observed used concrete objects to illustrate abstract math concepts within the observed lessons. The concrete objects included: fraction rulers to teach fractions; geoboard to teach perimeter and area; Kleenex and cereal box to teach surface area; wooden geometric shapes to discuss geometric vocabulary; and student-made rectangular prisms to teach surface area and volume.  The teachers who did not use concrete objects were often engaged in lessons in which manipulatives were not appropriate such as data collection and test review.  The observed use of concrete objects to facilitate the understanding of abstract concepts was present in both years of observation.

Sixty-six percent of the candidates used some variation of group work of their math lessons such as peer-to-peer tutoring, cooperative groups, think-pair-share and jigsawing.
Candidates also participated in interviews at the end of the classroom observation period. All candidates reported that they felt well-prepared to guide students’ learning of the lesson content. Candidates agreed the Middle School Math program resulted in the candidate asking higher-order questions and encouraging students to actively think and question the math material independent of teacher inquiries.

Second year observations evidenced an increase blending of science and math content during instruction. One of the candidates is also the grade’s science teacher and she feels as though she is often using concepts from one subject and relating it to the other; for example, graphing is important and relevant in both science and math and she can relate one lesson to another previous lesson. Another candidate described a math project completed by her students in which they used the scientific method to create a data analysis project in math class. One candidate described the school’s science curriculum as inquiry-based; therefore, she felt it was easy to apply math concepts in the science class. For example, her students measured the evaporation levels of water in various locations within the classroom.

In the dozens of classroom observations only three instances of technology use was observed.  As part of the universities NCATE evaluation the instructor of the math content courses were asked to rate the candidates’ proficiency with technology.  She rated 85% to be Target, 15 % to be Acceptable and 0% to be Unacceptable.  This data supports the pedagogical questionnaire data there was a statistically significant increase for the statement: Use of technology in support of mathematics instruction (p =. 041)

Comments made during focus group also substantiated that the program had impact on the candidates’ practices. “ I moved away from teaching isolated concepts; [started] making connections between different subject matter. [Math is] not the algorithm, allow [students] to develop an understanding.” I am able to give kids multiple strategies for different type of thinkers; here’s five different ways to solve this one problem.” “[I am] more reflective, figure out what is working in classroom and what is not. Through [program] coursework [I have] been able to evaluate it, change it and then reevaluate it to see if it worked.” “As a bilingual educator, all of the courses in the program made me more conscious of bringing different strategies to address learning needs of specific students rather than ‘one size fits all’ teaching strategies.”

While not an original component of the evaluation, all candidates were required to complete a MEd culminating project of a portfolio consisting of a practicum curriculum project and reflections of their experience.  Eighty percent scored at the highest level on the rubric: Target and the other 20% scored Acceptable.  Their projects were filled with lessons that actively involved the learner and aimed at conceptual understanding.  Many projects gave students some control of their learning and encouraged reflection making the lessons metacognitively explicit.

Results from observations, interviews and the focus group provide significant evidence that the Middle School Math program impacted the pedagogical practices of the candidates. These results also provide evidence of the program’s success related to the professional development of math teachers because the program changed how teachers taught, and these new methods are often associated with student achievement.

**DISTAL VERSUS PROXIMAL OUTCOMES**

Funding agencies often expect grant evaluations to focus on students’ achievement gains.  This is a logical consequence of the national directive to improve student achievement, not to mention an easily understood measure for the legislators who ultimately vote to continue funding.  While the desired outcome of this grant is to affect student learning, the use of student achievement measurements to determine program success is difficult and can often be misleading when determining program value.

When determining the value of a program using outcomes, it is essential that the outcomes reflect the program design and intentions. Distal outcomes such as improved student achievement is the result of more proximal outcomes; that is, changes in teachers’ knowledge, pedagogical perceptions and implementation of new skills in the classroom. As the outcome measures move further away from the immediate effect of program activities, it becomes more difficult to attribute these distal outcomes to the program.

Due to the requirement of the funding agency, the grant evaluation plan included a quasi-experiment using the Illinois Goals Achievement Test (IGAP) as the measure of student achievement. While this was approved, it became apparent to the grant team that the IGAP was an inadequate measure for a variety of reasons.  The IGAP does not report math sub scores; therefore, the scores could not be linked to the specific content courses in which the grant teachers participated.  There was a high year-to-year transitory rate for the LEA teachers who participated in the grant and that made a year-to-year comparison of scores less valid.  Even for teachers who remained at the same school teaching the same grade for consecutive years the scores had questionable validity.  Data requests to the LEA were restrictive.  It was impossible to get teachers’ individual student scores from the previous year to compare to the scores earned in the current year. The evaluation team had access only to the teachers’ consecutive year aggregate scores, which limited the accuracy of student achievement gains, and the ability to establish any kind of cause and effect relationship between student achievement and the program.  Our inability to conclude that student achievement gains, or lack thereof, were attributed to the program was perplexing because the data from the proximal measures of teacher content knowledge and change in practice and pedagogy were encouraging.  The question arose as to what measures were most reflective of the program’s success.  To investigate this query, a formal logic model was developed.

The formal logic model provided a map of how strategies connected to long and short term outcome measures.  Furthermore, the model detailed the interrelationships of the partners and identified boundaries, natural or imposed, that limited the measures.  This enabled the evaluation team to more closely appraise the grant activities relative to their outcome measures. It also forced the team to identify necessary preconditions that lead to increased student achievement. It was this identification of preconditions, and the boundaries imposed on those preconditions, that led the grant team to determine the outcome measures most reflective of program success were measures of teacher content knowledge and pedagogical practice.

The program participants were readily accessible for pre and post testing in the six mathematics content courses taken as part of their program.  Content knowledge growth was measured using Diagnostic Teacher Assessments in Mathematics and Science (DTAMS) assessments in five of the six math content courses.  This was facilitated by a productive collaborative relationship between the grant administration in the School of Education, and the mathematics faculty in the College of Arts and Sciences.

Teacher's pedagogy perspective was assessed first in an education class during the second semester of the grant and again during the last semester of the grant using an on-line self-administered survey based on Classroom Mathematics Teachers Questionnaire, developed by the Horizon Research Inc. One-hundred percent participation was attained because of the ready access the professor, who was also grant director, had to the candidates in these two classes.  It was evident from the initial administration of these assessments, and the information provided from the logic model, that these tools alone did not give the depth of understanding desired regarding content knowledge and pedagogy.   Not only was the depth of understanding a concern but also the realization that the evaluation had no way to determine if the changes in content knowledge and pedagogy had made their way to the teachers’ classrooms. This realization highlighted the fact that the qualitative data (i.e. observations and teacher interviews) would be essential for determining the transference of knowledge and skills into practice.

In the second year of the program, the budget was modified to make possible the hiring of a graduate student to perform classroom observations, teacher interviews and data analysis.  The student selected, a Ph.D. candidate in school psychology, was familiar with educational concepts and had experience conducting qualitative observations and interviews.  The graduate assistant received professional development with the Inside the Classroom Observation and Analytic Protocol instrument. Upon successful completion of the PD, the graduate assistant observed the program participants’ mathematics classes. After each observation, the teachers participated in a 30-minute interview.  The analysis of this qualitative data provided evidence that the increase in content knowledge revealed by the DTAMS had, in fact, made its way into the classroom.

In addition, the program participants were implementing research-based instructional methods at a significantly higher rate. Through this triangulation of data collection, the grant could corroborate findings and strengthen the evidence of the positive effect the grant-supported program was having on teacher content knowledge and pedagogical practice. However, these encouraging findings were not supported by the ISAT data quantitatively analyzed relative to student achievement gains.  Achievement data findings were inconsistent but in general failed to uncover statistically significant gains in student learning.

The logic model was also valuable in helping the grant administrators understand the dynamics and boundaries of measuring student achievement.  The interrelationship between the grant partners at the university with those in the central administration of the LEA never developed to the point of working collaboratively toward more effective measures.   The original contact person in the district administration, who helped write the grant proposal, had transferred and the position remained unfilled.  Attempts to forge relationships with other members of the central administration were unsuccessful due to frequent changes in administration and the perspective on the part of CPS that this was a provider –consumer relationship instead of a partnership.   This caused problems in gaining access to the data that would have best measured student achievement.  Tracking individual students became impossible and only aggregate classroom data was made available.  These data lacked the power to uncover advances in student achievement.  The logic model again prompted the grant administration to gather qualitative data through observation and interview.

The analysis of observation data revealed that students in the grant teachers’ classes were more engaged in learning through active participation, peer group work, inquiry based, project based and problem solving learning, peer group.  The teacher interviews substantiated this conclusion. One can infer that with increases in engagement come increases in student learning and achievement.

While grant administrators were confident the program had produced increases in teacher knowledge and improved pedagogical practice, the impact on student learning was not clear: quantitative data indicated no increase but qualitative data provided evidence of increases.  To facilitate resolve of this conundrum the evaluation team decided to convene focus groups with the candidates in the program.  The first focus group turned out to be more revealing than anticipated.  The participants’ open and frank conversation relative to all aspects of the program outcomes supported evaluation findings regarding the increase content knowledge, pedagogical perspectives, and pedagogical practice. Program candidates even provided statements about the program’s impact of increased student achievement.  Some members of the cohort had kept track of individual student achievement from year to year and reported significant gains.  One participant reported having the highest scores in the school the past two years; she attributed this feat directly to the affect of program on her advances in knowledge, method and confidence.  The district denied request to form student focus groups to gather data on student learning due to confidentiality concerns, but candidates were adamant that the students would have reported an increased interest, engagement and achievement in mathematics over previous years.   It was through these observations, interviews and focus groups that the grant administrators became confident the program had had an impact on student learning and achievement.

**CONCLUSIONS**

While increased student achievement is the ultimate goal of teacher professional development programs, the sole use of student achievement scores as an indicator of program success may only misinform program developers and funders as to the value of a professional development program.  In this evaluation, results of the student achievement measures provided mixed information of the program’s value. In addition the limitations (e.g. small sample size, changes in students from year to year, and inability to control for impact confounding variables on students’ achievement scores) associated with this data decrease its contribution in generating meaningful information to determine program value and taken by themselves may have led to the conclusion that the program was unsuccessful.

By examining more proximal measures, primary outcomes (increased math content knowledge, increased math pedagogical knowledge) and intermediate outcomes (changes in pedagogical practices in the classroom), the evaluation team was able to measure the program’s value more accurately.  By evaluating changes in DTAMS scores and shifts in pedagogical perspectives questionnaire, the evaluators were able to establish the program success at meeting the primary goals of increasing teacher’s content knowledge and pedagogical knowledge.   While the quasi-experimental approaches used to assess these outcomes served to substantiate the program’s affect on content knowledge and pedagogical perspectives, it did not provide any evidence that this new knowledge was being used in practice or how it affected the participants personally.

Classroom observations evidenced that participants were transferring their new content and pedagogical knowledge into the classroom while interviews provided evidence that the changes in the classroom practices were linked to the learning occurring in the professional development program.  The focus group data provided a more global understanding of the program’s influence on the participants’ confidence, their beliefs about math instruction, their classroom practices and the impact the program has had on their students’ learning. Furthermore, the use of observations, interviews and focus groups provided some of the more descriptive and informative data regarding the program’s outcomes, as well as changes that could be made to improve the program.