Final Student Outcomes Study for the Gadsden Mathematics Initiative (GMI)

To: Paola Sztajn, NSF Program Officer, psztajn@nsf.gov

To: Joy Frechtling at WestEd, mailto:joyfrechtling@westat

To: Wanda Guzman, Sheila Raihl, & Yvonne Lozano, Gadsden Independent School District

From: Karin Wiburg, Associate Dean for Research, New Mexico State University

Attached is the final report on the Gadsden Mathematics Initiative *Student Outcomes Study*. An overview of previous reports submitted is also included below. The final report integrates findings related to the GMI and its effect on student achievement from 2003-2006.

1. The *first report* was an overview of the state of the research in February, 2003, as requested by the National Science Foundation.

2. The *second report* was a much more extensive overview of the results of the 2003 data collection of student test scores and teacher observations, provided to West-Ed & NSF in July 2004.

3. As a result of the second report, changes were made in the research design and a partnership was formed with Dr. Marta Remmenga, Professor of Statistics at New Mexico State University and her graduate students. We discovered after the first year that the research required a new form of statistical analysis using a hierarchical linear model, a mixed effects model, which could control for some of the variables present in school research, such as the differences between schools and classrooms, when considering the effect of professional development and student achievement. The *mid-point evaluation report, 2005*, used this Mixed Effects model. The 2003-2005 report was extensive and included two parts. *Part A* was a general overview of the results and *Part B* provided detailed information on the development and use of a new mixed effects model

4. Many people worked on the Student Outcomes Study including:

Dr. Karin Wiburg, Associate Dean, College of Education, New Mexico State University

Dr. Marta Remmenga, Professor, Statistics, New Mexico State University

Dr. Dennis Clason, Associate Professor, Statistics, NMSU

Erika Glaser and Yi Hu, statistics graduate students

Ken Korn, Director of Evaluation for the Gadsden Independent School District

Dr. Cathy Kinzer, Project Coordinator, Mathematically Connected Communities, Math Educator

Numerous observations were also done by observers trained by Horizon Research.

Gadsden Mathematics Initiative - Final Student Outcomes Study 2003-2006

The GMI (Gadsden Mathematics Initiative) is a K-8 district wide effort to enhance teacher content knowledge of mathematics, increase teachers' pedagogical skills, and gradually implement standards-based resources (*Investigations* and *Connected Mathematics*) at the elementary and middle grades. Funded by the National Science Foundation, as one of the last Local Systemic Change Initiatives (LSCI) in 2000, the GMI funding was completed in December 2006 although GMI continues in the district. The project was a partnership between the Gadsden Independent School District (GISD) and New Mexico State University (NMSU). The GMI was a district-wide effort with top-down and bottom-up support that provided up to 40 hours/year of intense professional development to teachers, through both summer academies and follow-up workshops. Initial professional development was provided district-wide. As the project, teachers and principals implemented this project, the focus for professional development changed from district-wide large trainings to school-based professional development communities and classroom mentoring facilitated by mathematics coaches, called in Gadsden, Mathematics Process Trainers (MPTs). During the period from 2000-2006, the district took increased financial and administrative responsibility for the GMI, slowly moving the project under district operational funding.

The GMI project is continuing with an emphasis on high-quality mathematics learning for teachers and the implementation of a rich and challenging standards-based mathematics curriculum for all students, K-12. Mathematics coaches have been hired from district funds in all K-8 schools and the district is currently hiring MPTs for each of the 3 high schools and introducing intensive professional development for 9-12 teachers. In addition, 17 of the top mathematics teachers in GISD are currently completing a masters program in Teaching Mathematics that was developed by NMSU mathematicians and educators in collaboration with GISD. The district paid top mathematics teachers to complete this two-year plus two-summer program. Strong relationships continue between GISD and the NMSU mathematics initiative and GISD is now a partner in the new state MSP, Mathematically-Connected Communities (MC^2).

Current State of Mathematics Achievement in the Gadsden Independent School District

An overview of the current state of student mathematics achievement in GISD is presented below. This is followed by a discussion of the Student Outcomes Study which was conducted at the request of the National Science Foundation to explore possible relationships between the professional development provided and student achievement in a 95% Hispanic, low-income district with over 60% of the students in the district classified as English Language Learners.

The proportion of GISD students who scored at or above the proficient level at the beginning and at the end of the project is shown below in Figures 1 and II. In 2000, only the section of the traditional norm-referenced state Terra Nova on problem solving was used for assessing student achievement, not the entire test, because this section aligned most closely with the later criterion-referenced tests provided in 2004-2006.

Figure 1 – GISD District Student Math Proficiencies at the beginning of the project

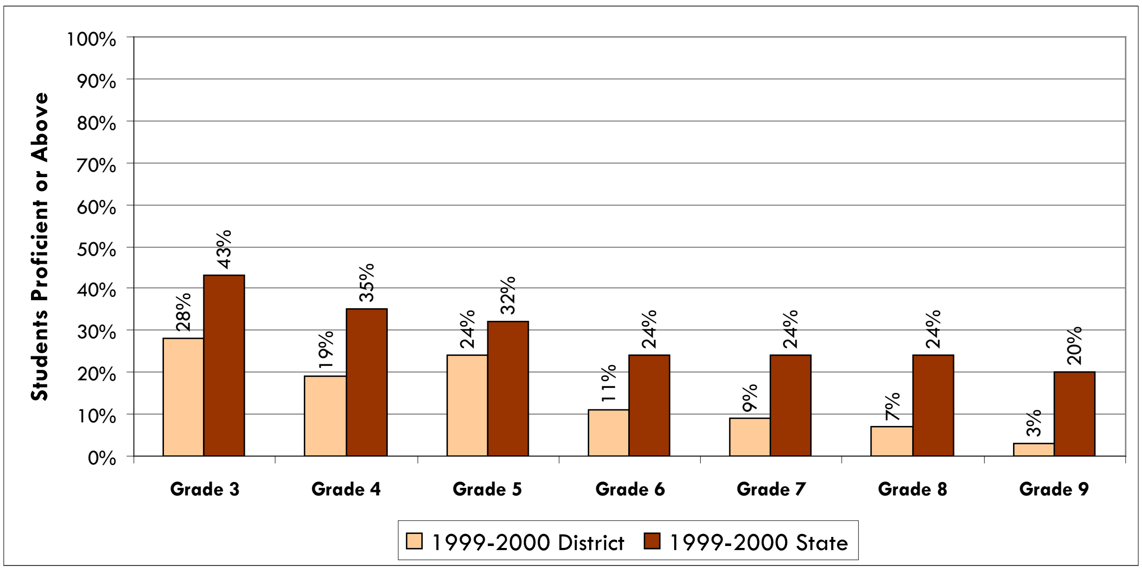
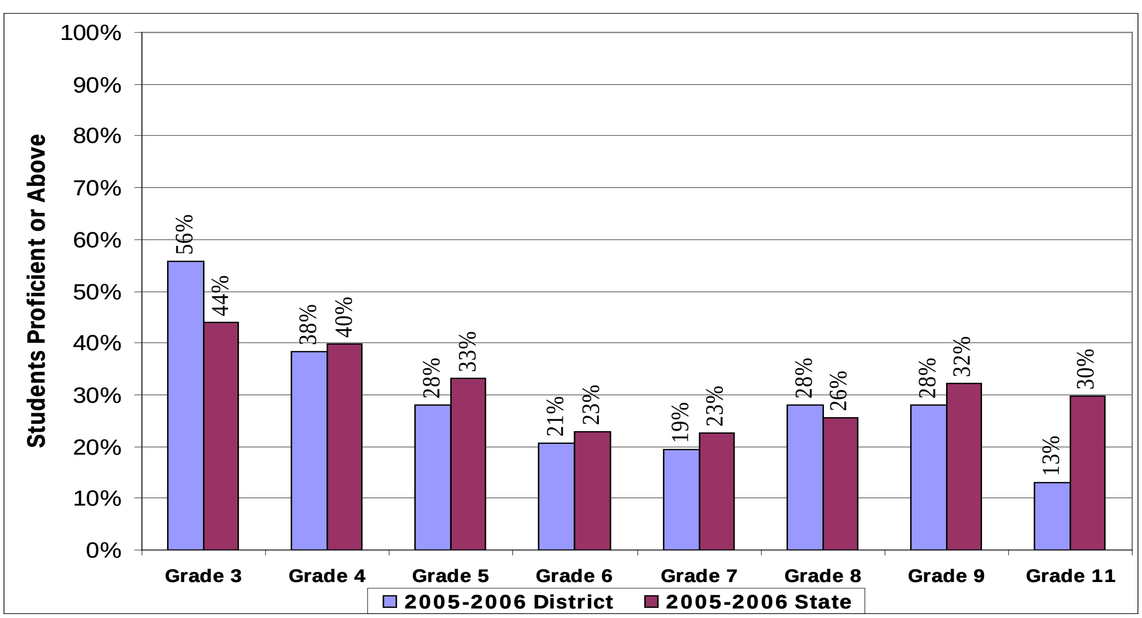


Figure 2 below shows the current performance of GISD mathematics students at the end of the NSF-funded Gadsden Mathematics Initiative.

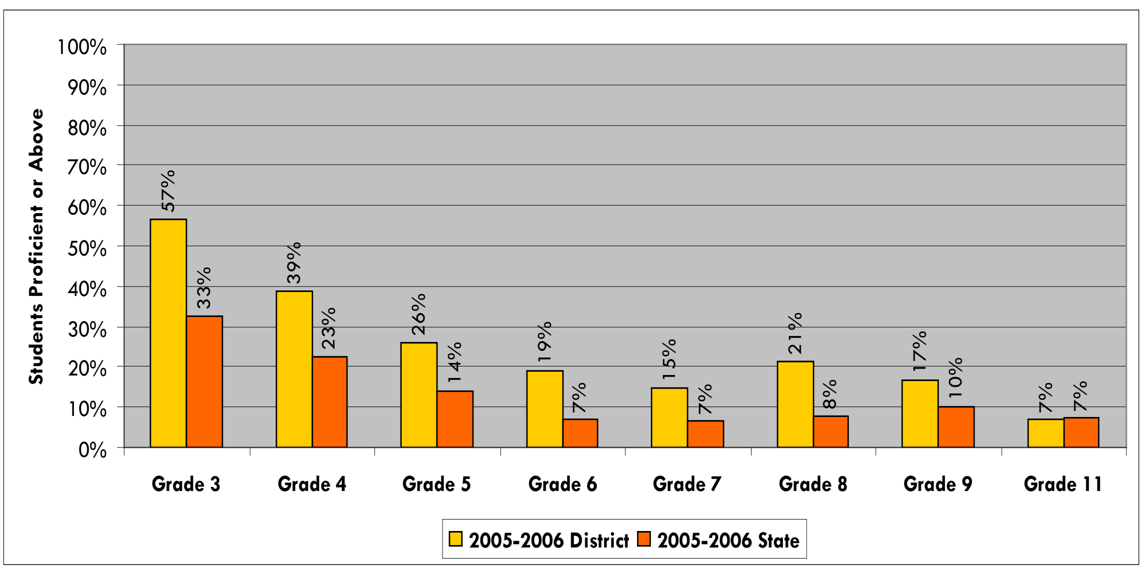
Figure 2 – GISD District Student Math Proficiencies at the end of the project



In Figure 2 you will notice that the students in grades 3 and 8 showed the highest increase in percentages of competencies in relationship to state average scores. These were the grade levels in which students had participated for the longest period in the GMI (K-3, and 4-8). These grade levels continued to show more gains when the sub-groups were also examined. This data provides a larger view of the entire GISD population that what was available by just reporting the GMI Student Outcomes Study (SOS). The SOS was limited to only 8th and 4th grade for reasons related to the availability of test scores at only those grades.

Because of a state and national concern regarding the achievement gap in mathematics and science achievement between students who are learning English and mainstream students, the growth in mathematics proficiency for ELL students in GISD is particularly important. The level of mathematically proficient students at GISD is greater than the state in all grades except the 11th (which was not included in the GMI initiative). An administrator at the state department suggested recently that "If you are an English Language Learner in New Mexico it seems that the best place for you to go to school is in Gadsden."

Figure 3. Proportion of 8th grade English Language Learners at or above proficiency levels on the mathematics section of the NM-SBA test in the Gadsden Independent School District and in state of New Mexico, 2006.



As can be seen by this data on all students tested in GISD and the state, 3rd-11th, the GMI had a large impact on a traditionally low performing district and to a large extent closed the achievement gap. In fact GISD was number 88 out of 89 districts as ranked in student test results at the beginning of this project (2000) and has moved to x in the rankings of district achievement scores.

The GMI Student Outcomes Study

In 2002, the National Science Foundation (NSF) asked the Principle Investigators of the GMI to begin to study possible relationships between teacher professional development and student achievement. In response to that request, a study was designed to address some of the possible reasons for student success in mathematics in the Gadsden school district. The following section provides a short overview of the Student Outcomes Study and changes made during the study in response to our yearly analysis. This is followed by a discussion of the population studied, the research questions, methodologies used to address these questions, findings, conclusions and suggestions for additional research.

The GMI student outcomes study (SOS) began in 2003 and continued through the end of 2006. It was focused on fourth and eighth grade teachers and their students, because for the first two years of the study, the new criterion-referenced testing was piloted only in 4^{th} and 8^{th} grades. All K-8 teachers in the district received professional development training on the use of the new NSF-approved curriculum materials that would be used, Investigations (K-5) and Connected Mathematics (6-8). They also received instruction in curriculum alignment with standards and assessment, instructional strategies for teaching an inquiry-based curriculum, and in some cases, Lesson Study training. Lesson Study is a collaborative process where teachers devise a lesson, teach and observe the lesson amongst his/her peers and re-teach the lesson in an iterative cycle of professional learning (See Wiburg & Brown, 2007 for a discussion of implementation of Lesson Study in the U.S.). Professional development was offered during summer academies with follow-up training and coaching available during the school year.

While gains in student achievement can be seen district-wide, the GMI student outcomes study (SOS) was designed to explore what aspects of the implementation of the GMI professional development initiative might have had the most effect on the professional growth of teachers and in turn, how teacher behaviors in the mathematics classroom might affect the mathematics achievement of their students. The SOS focused on the following research questions.

Research Questions

1) Does the level of implementation of the GMI professional development (as measured by Levels of Use (LOU) and Classroom Lesson Observation (CLO) instruments) affect the level of student achievement as measured on the state standardized criterion-referenced and norm referenced tests administered during the period of the GMI project?
2) What do we know about the changing nature of pedagogy in classrooms as a result of the GMI?

   2a) What do classroom observations tell us about classroom instruction, student engagement and type of learning, and how students are learning mathematics?

   2b) What, if any, relationships, exist between the level of teacher use of Professional Development (as measured on the Levels of Use scale described under Instrumentation) and factors related to teacher classroom practice (as measured on the Classroom Lesson Observation described under Instrumentation)?
3) Does the type of professional development (Lesson Study or Other), or the amount of professional development affect the level of student achievement as measured on the standardized annual mathematics tests administered by the state?

<u>Targeted Schools and Populations</u>

The GMI Student Outcomes Study was carried out in the Gadsden Independent School District (GISD). The student population in GISD remained steady at approximately 94% Hispanic during the period of the study. At least one-third of these students are Spanish language dominant when they enroll. An estimated 60% of the students in the district come from homes in which the dominant home language spoken is Spanish, not English. With respect to socio-economic status in the district, approximately 85% of families are classified as low income or poverty status and the district is a 100% free lunch district. The study focused specifically on teachers of 4^{th} and 8^{th} grades and their students, since these were the only grade levels that would be continuously tested from 2002-2006. Fourth and eighth grade served as the pilot groups for taking the new criterion-referenced tests that were being developed in the state of New Mexico. The study included all students, with the exception of students who were in the lowest level of English classes (Spanish-speaking newcomers), or students who were in self-contained special education classrooms. However, Spanish was spoken in many of the classrooms we observed, even those not designated as bilingual or transitional. All bilingual transition classrooms and classrooms in dual language English-Spanish programs were included. Classrooms for newcomer monolingual Spanish-speaking students were not included since the focus of these classrooms was to teach English, rather than grade-level mathematics.

Instrumentation

Teacher Observation. All teachers were observed using a standardized classroom observation tool, the *Classroom Lesson Observation* (CLO) which was adapted from the Horizon Assessment Observation Protocols for evaluating systemic change initiatives. Using the *Classroom Lesson Observation* (CLO) Worksheet, the trained observer is asked to state the strengths and weaknesses in each of the five components of a lesson, including an overall rating of the lesson. A second instrument, the *Levels of Use* (LOU) scale, was developed by the GMI P.I.'s, Karin Wiburg and Wanda Guzman, using the Concerns Based Adoption Model (CBAM) approach developed by the Southwest Education Development Laboratory (SEDL). The Levels of Use scale was designed specifically for the Gadsden Mathematics Initiative based on the Southwest Educational Development Labs (SEDL) guidelines for the design of a Concerns-based Assessment (CBAM) that could measure the effect of an innovation on teachers. The scale was designed to measure the level of implementation in the classroom of the professional development provided by the GMI, and reflects specific goals of the GMI professional development including the depth of the mathematics instruction, how much it contributed to student-generated discussions of mathematics, the number and type of engagement of the students, and the classroom environment in terms of student participation. The *Levels of Use* scale contains fourteen items rated on a 0 to 6 scale.

A final observational tool, the *Classroom Snapshot*, was based on the protocol used by NMSU's Collaborative for Teacher Education Programs (CETP) to observe classrooms of pre-service teachers as the university and later to observe these same students once they became teachers in classroom. This snapshot was a slightly shortened version of the Horizon Protocol for observations in classrooms involved in many of the NSF systemic change initiatives.

As much as possible, evaluators trained by Horizon at the University of North Carolina served as observers and filled out the LOU, the CLO, and the Classroom Snapshot observational forms during teacher observations of all available fourth and eighth grade teachers in the district every Spring. (All teachers were observed except teachers of self-contained special education classrooms and teachers who were teaching the lowest level of English to newcomer students).

Observers indicated on the Classroom Snapshot the type of instruction, the engagement of the students and the type of learning in which students were engaged receptive, structured application, or construction of understanding and in terms of level of engagement, high, medium or low. For students to be rated as highly engaged, at least 80% of the students has to be seen as highly engaged during the observation.

Observers looked at the classroom every five minutes and noted each of these variables and filled in a time chart as seen below: As described below the data was analyzed as separate events which were noted every 5 minutes. Because of the number of teachers observed, and the lack of additional funding in the grant for hiring high-level researchers, the primary evaluators found it necessary to also train several doctoral students. Every attempt was made to develop inter-rater reliability by new observers doing observations with the trained evaluators. Refer to Appendix 1 for instruments used in observations.

Changing Measures of Student Achievement in New Mexico. During the GMI, the state of New Mexico, like most states between 2002 and 2005 made changes in how they measured student achievement on their yearly annual evaluation. As mentioned in conjunction with a presentation of achievement at the beginning and end of the project, the state moved from using an annual state norm-referenced standardized test, the Terra Nova, to a newly mandated criterion-referenced test (CRT). This test was designed by CTBS McGraw Hill and given to students in grades 4 and 8, beginning in 2003. Students in GISD also took the Terra Nova norm-referenced test in 2003. In 2004, students took only the CRT. During 2004, the NM State Department of Education decided to develop a new criterion-referenced test with a new company, Scott Foreman. This new test is called the New Mexico Standards-Based Assessment (NM -SBA). It is similar to the earlier CRT, although a larger percentage of points (50%) are given in response to open-ended constructed response items, both short and long. This new test has been highly-rated because it is one of the few state tests which aligns with the National Assessment of Educational Progress (NAEP), a national measure of students' understanding of mathematics concepts. It asks students to explain their mathematical thinking and demonstrate in writing and using charts and graphs their understanding of mathematical concepts. While aligned with the standards, the new test was not aligned with the traditional mathematics teaching used in many schools which asked students to memorize and perform procedures. On the other hand, the new criterion-referenced tests were well aligned with the goals of the Gadsden Mathematics Initiative which emphasized inquiry learning and reasoning for all students.

Data from the teacher observations instruments, from pre- and post- interviews of teachers and from these multiple yearly state standardized measures of student achievement were used to explore possible relationships between levels of implementation of the professional development, classroom lessons, types and hours of professional development and student achievement.

Data Collection

Teacher data was collected in the fall of 2003 through use of the CLO, the Classroom Snapshot and LOU instruments. Attempts to connect this data to 2003 student achievement data was hampered because teachers had moved classrooms between the Spring and Fall terms and the sample of teachers for whom we had matching students was small. Beginning in 2004, teacher observations were performed in the spring during the two months after the March standardized testing period to facilitate matching student and teacher data. The 2004, 2005, and 2006 observations of classrooms and student testing were all done in the spring, although student test results were typically not available from the state until the following October or November.

Observers who completed the observation instruments were primarily trained facilitators through the Horizon training at University of North Carolina. However, due to a lack of funding for hiring high quality observers, the main researchers on this project worked to train their graduate students in how to do the observations. This was done by pairing new observer with experienced observers until there was at least a 90% level of inter-rater reliability.

Pre and Post Interviewing. Teachers were asked at least a week before a planned visit to accept observers in their classrooms. They were given a pre-interview over the phone during a free period prior to the observation and were asked to fill out some additional information related to the lesson the observer would see. Observers then came and watched the lesson and the classroom for 45 minutes to one hour and filled out all the observation forms. As much as possible the observers tried to find a few minutes during recess and/or lunch or right after school, to ask the teacher how they felt the lesson had gone and what their students had learned. Teachers were also asked in which areas the students might need additional help. During one year (2005) we decided to try to ask them more about their mathematics backgrounds but this process was not effective, it was too much at one time. We had hoped to do some additional separate interviews related to the teachers' math backgrounds and knowledge but we simply didn't have the human resources in this study to do all we wanted to do.

Student achievement data was collected on the standardized norm-referenced test (using the scores from the problem-solving Terra Nova in 2003) and then using the two different standards-based criterion-referenced tests for all fourth and eighth grade students every year after that. In 2003, there was a criterion-referenced test, but scores were only reported to districts by students and not connected to teachers. Therefore in 2003 we used the Terra-Nova scores, which were reported by teachers and focused on GMI-type problem-solving type questions which were related to the types of questions on the new state criterion-referenced tests. In 2004, 2005 and 2006, students took only the criterion-referenced tests and student scores were provided to the research team by teachers, for all elementary schools. This made it possible to use a statistical model to connect teacher's professional development with student achievement scores. In middle school, the student scores were presented by home room, not by mathematics class, so we got class roosters for math classes and connected the middle school students in the appropriate grade level math class with the math teacher we had observed. We were assisted in the process by Ken Korn, the district head of evaluation in GISD.

Statistical Methodology

In the first year of the study (2003), a factorial analysis of variance model was used to examine the effects of the GMI project on teachers and student achievement measured as classroom averages. Factors in this model included mean rank from the LOU instrument and overall rating (on a six point scale) from the CLO as well as their interaction. Correlations between the CLO and LOU instruments were also estimated. In addition, results were examined for the 4th and 8th grade classrooms combined and included self-contained special education classrooms.

In 2004, the statistical model and analysis was redesigned with the assistance of a statistician and her graduate student, Erika Glaser. The new analysis uses a mixed effects statistical model and takes into account the nested nature of students within classrooms and teachers within schools. Variability among schools and among teachers within schools is accounted for by this new analysis as explained by Glaser:

> *…a mixed effects model might be an appropriate choice when analyzing individuals nested within natural occurring hierarchies, like students within classes and classes within different schools. A mixed effects model can be used to examine the behavior of a low level outcome, such as student scores, as a function of higher order predictors, such as classroom and school. A mixed effects model can also be used to handle situations where there are unequal variances or a lack of independence. By properly specifying the mixed model, a variety of specific instances of multilevel models may be analyzed thus resulting in a more suitable statistical analysis.* (Glaser, 2004)

The new analysis required that data include individual student scores rather than classroom averages. The mixed effect model was applied to the 2004, 2005 and 2006 data. A description of how the models were fit to the data each year is provided in Appendix B. All statistical analyses were performed using the BASE, GLM and MIXED procedures of SAS (Statistical Applications Software) version 9.0.

It was also determined after the first year of the study, that 4th and 8th grade teachers of mathematics were two very different populations and grouping them for the study was a mistake. They had very different tasks and responsibilities and their students were involved in different kinds of mathematics activities, thus the two grades were analyzed separately during the last three years of the initiative. A resulting problem with this was that there were a very small number of 8th grade teachers making it difficult to find statistically significant results as a result of the numbers. However, we believe that it was the right decision to look at 4th and 8th grade results separately and some of the results were statistically different based on the grade levels themselves, not just the number of subjects. Because the students in self-contained special education classrooms were working at very different levels than the average 4th or 8th grade classrooms and because of the limited number of highly-qualified evaluators and observers, these classrooms were excluded. However, all special education students who were in mainstream classrooms were included in the study.

Observers were asked to used the Classroom Snapshot observation tool to describe the classroom and lesson and then to fill out codes for instructional events, types of learning, engagement of students and the use of tools for learning every five minutes while observing in the chart shown in Table 1 below. The following categories were used to code instructional events. .

Table 1: Codes and Chart used to Record Instructional Events

Lec = Lecture	IS = Independent Seatwork
LecQ = Lecture with Questioning/Discussion	SP = Student Presentation
CD = Class discussion	SGD = Small group work
PM = Problem-modeling	TIS = Teacher interacting with Students
I = Interruption	A = Assessment

Chart used to record Instructional Events every Five Minutes:

	5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	5055	55-60
Instruction												
Student Engagement												
Type/Learning												
Tools												

A new category was developed for a type of Instruction during the GMI that had not been included in the original instrument. There was a type of teacher behavior that emerged which was not a lecture, a lot like leading a classroom discussion, but also included the use of questioning as an instructional strategy for helping students describe mathematical ideas in multiple ways and share strategies for number computations and operations. The research team came up with a new category of instruction that integrated classroom discussion, teacher lecture-like control and the use of instructional questioning. We called this category *Lectures with Question* or *LecQ*.

Teachers had learned to not immediately tell students the answers but to increase the students' active thinking and learning time by asking good guiding questions. The section of the Classroom Snapshot in which observers made a sketch of the room also indicated an increase in the amount of classrooms in which students were seated at small tables as the project continued. The amount of time spent on student small-group work and student-student communication was supported by the fact that only a few classrooms had students seated in traditional rows by the end of the project observations. Most children sat at 4-5 person tables or at least were paired with another student in the seating arrangement.

The Limitations of Educational Research Methodology

A major reform effort in government has led to demands of "scientifically-based research" as a result of the *No Child Left Behind Act* and the reauthorization and reorganization of the Office of Educational Research and Improvement as the Institute for Educational Science (Maxwell, 2004). Government funding agencies are asking for research based on quantitative designs and methodologies for studying what effect reform moves are having on schools and student achievement. Most educational research directed toward instructional and school improvement tends to involve implementation of a plan or strategy designed to solve a specific problem in our educational system. To quantify the effectiveness of such plans is a challenging problem.

Determining how the results of teacher training might effect student achievement is especially difficult. There are limitations in designing studies dealing with teacher development because of uncontrollable factors affecting the implementation of the teacher training. After a teacher has received training, there is a lag before we would expect to observe student improvement. To handle this, educational research dealing with effects of teacher training on student outcomes is conducted over a period of at least three years. However, due to high teacher turnover, educational researchers often end up with fewer subjects than originally planned. The GMI extended over five years and although teacher turn-over declined dramatically in the GISD, with only 35 teachers replaced in 2006 compared to around 100 in 2003, some teachers moved to different schools within the district or changed grade levels within the same school. In addition, the district changed two K-3 schools into K-6 schools. These changes were particularly problematic because only grades 4 and 8 were included in the study. As a result, between 40 and 50 teachers were observed each year on average, but only a subset of 10-12 teachers were observed every year of the study. See Table 2 below which shows the numbers of teacher observed.

Table 2. Number of teachers observed at each school in the Gadsden Independent School District each year of the Gadsden Math Initiative.

School	2003	2004	2005	2006
Chaparral M.S.	3	3	2	3
Gadsden M.S.	4	4	3	6
Santa Teresa M.S.	3	2	3	4
Berino E.S.	6	6	2	5
Desert Trail E.S.	9	7	4	2
Desert View E.S.	4	3	3	4
La Mesa E.S.*	0	3	3	
La Union E.S.	7	3	2	2
Santa Teresa E.S.[1]		3	0	2
Loma Linda E.S.	7	6	5	3
Mesquite E.S.	0	3	0	2
Riverside E.S.	0	4	4	0
San Miguel E.S.	3	3	3	3
Sunland Park E.S.	2	2	3	2
Sunrise E.S.[1]			3	3
Vado E.S.[1]				2
Total	48	52	40	43

[1] Schools were newly built in the first year showing observations. * School was closed in at the end of 2005.

Another common problem in educational research is the inability to randomize teacher professional development treatments within a school or district. To implement an ideal designed experiment, equal proportions of the teachers within a particular school or district should receive each level of the treatments (including the control treatment). This is rarely implemented in practice because it would be difficult for a principal or superintendent to justify providing different types or levels of professional development to teachers in the same school or even in the same district. Even if a randomized design could be implemented, lack of independence between treatment groups would soon develop because teachers in the same school tend to share materials and talk to each other about teaching activities (Kramarski & Merarch, 2003). Because of these design limitations, it is difficult to attribute changes in student scores to the treatment factors alone. In addition to the planned treatment factors, student scores may also be affected by the demographics of the population of students served by the school, the level of teacher support provided by the school principal, the district's support of teachers and schools (including extra paid training) and other types of teacher development offered in the district.

Historically, teacher effectiveness has been measured by self-evaluation and observations of teachers. Currently there is an effort to measure the effectiveness of teacher training in terms of what kinds of professional development result in student learning gains (Hill, Rowan, & Ball, 2005). The GMI Student Outcomes study was designed to address this increasingly important research area regarding how professional development of teachers affects gains in both teacher gains in pedagogical content knowledge of mathematics and student achievement.

In addition to the limitations mentioned above, other challenges were present in the GMI. There were frequent changes in state-standardized tests in New Mexico during the period of this study. Three different state-required annual standardized test instruments were used to measure student achievement. There were also changes in the school board and the removal of top-level administrators during the 2005-2006 school year. It is a tribute to the GMI program that during this year (the fourth year of the initiative) the schools continued the math reform and student scores remained high or increased even though the administrators who had originated the project were absent. Another challenge in this study was the limited number of highly qualified evaluators available to complete accurate classroom observations due to limited funding the research.

Disruption in the GISD District Administration

An interesting event happened on the way to full implementation of the Gadsden Mathematics Initiative. The Gadsden district had been through profound reorganization from 2000-2005 in terms of a complete change in their mathematics program and a change from a traditional textbook reading program to a new balanced literacy program. There were also changes in the bilingual program, which in a few schools worked well, but in other schools was putting bilingual students in a situation where they were not getting adequate support for building literacy skills in either language. There have always been tensions in this district between the traditionally rural valley, from which most of the students come, and the many administrators and teachers who drive into the district from the neighboring larger cities of El Paso, Texas or Las Cruces, NM. For a number of reasons, in the 2005, a group of citizens from the valley began to organize a movement to get rid of the current superintendent, associate superintendent for instruction, the Director of Personnel and other top leaders, who were perceived as coming from outside of the district and perhaps making too many changes in the schools. There was a school board election in the spring of 2005 and the superintendent was removed. The associate superintendent, who had served as the P.I. for the Gadsden Initiative decided to retire and there was a great deal of turmoil in the top leadership in the district. This new school board, which was elected by a very small majority and soon seen as detrimental, as well as involved in some illegal issues, was recalled within 9 months and a new school board was put in place by the state. The superintendent returned to work as did other top staff.

A final challenge experienced in the GMI was the rapid growth in the numbers of teachers and students being added to the Gadsden district. At the beginning of this study, the Gadsden Independent School District (GISD) included 12 elementary schools and 3 middle schools. The district now has 14 elementary schools, 3 middle schools and will soon open a new high school. . GISD student enrollment grew from 13,100 students in 2000-2001 when the grant began, to 14,030 students today.. As the enrollment increased in GISD, so did the number of students who are English Language Learners (ELLs). According to the *New Mexico Accountability Data System*, there are now over 7,200 ELL students in the district. Students with learning disabilities (special education students) are also on the rise. Special Education documentation from the *New Mexico Data Accountability Data System* lists over 2,000 district students receiving special education services.

The increase in student enrollment and the addition of new teachers in the GISD, as well as the need to address the specific challenges of special needs students led the GMI leaders to request supplementary funding in 2005 to provide new teachers, ELL teachers and Special Education (SPED) teachers with GMI training. These supplementary funds were received and provided additional site-based professional development during the final stages of the GMI, including a focus on addressing the needs of special needs students and an integration of the new Algebra Strand implemented in the most recent New Mexico Mathematics Standards. In the 2006 student test data, GISD special education students scored above the state average scores in mathematics competency for all grades 1^{st}-6^{th}. Additional training was not done for SPED middle school teachers and the students at this grade level did not score above the state average.

Summary of Findings

The results of the data analysis are summarized in the following section. They are presented by year in some cases and in others are summarized over several years.

Research Question 1. *Does the level of implementation of the GMI professional development (as measured by the LOU and/or CLO instruments) affect the level of student achievement as* *measured on the criterion- referenced mathematics tests administered by the state in 4th and 8th grades?*

Research Question 1 - 2003

There was no detectable difference in student achievement for different observational ratings of teacher behavior. Residual diagnostics (examination of residuals from the statistical analysis) revealed severe unequal variance by school as well as an unaccounted for fixed school effect. In other words it made more of a difference which school the students were in, as compared to which teacher the students had. Specifically there was a significant difference in student performance by school but not by teacher.

In the Student Outcomes Study, we were interested in how professional development affected teachers, rather than schools, and how those teacher's teaching affected student achievement. The design for data analysis was changed with the help of Remmenga, a statistics professor at New Mexico State University, and a hierarchical data analysis mixed effects model was utilized that would account for differences between schools and between classrooms (See Statistical Methodology).

Research Question 1 - 2004

Using the new mixed effects model to account for variability between schools and within classrooms, no differences in student scores were detected in relationship to either the LOU or the CLO. The reason for the lack of discernable differences may be the result of the small number of teachers (especially in the eighth grade sample), possible differences between evaluators' ratings or some other factors that we had not identified. Residual diagnostics revealed that the variability among teachers within schools was adequately accounted for. It may also be that it takes a while for reform to affect student performance and it was too soon for changes in teacher knowledge or performance to affect student achievement.

Research Question 1 - 2005

Differences were detected in students' achievement as a function of the level of implementation of the professional development in the classroom as measured on the LOU scale in the fourth grade classrooms. For each one point increase in the LOU scale, students had a 3.7 increase in achievement. **Figure 4** below shows the relationship between the LOU measurements and student scores on the state standardized tests. No differences were detected in student achievement in relation to CLO measures for 4th grade teachers.

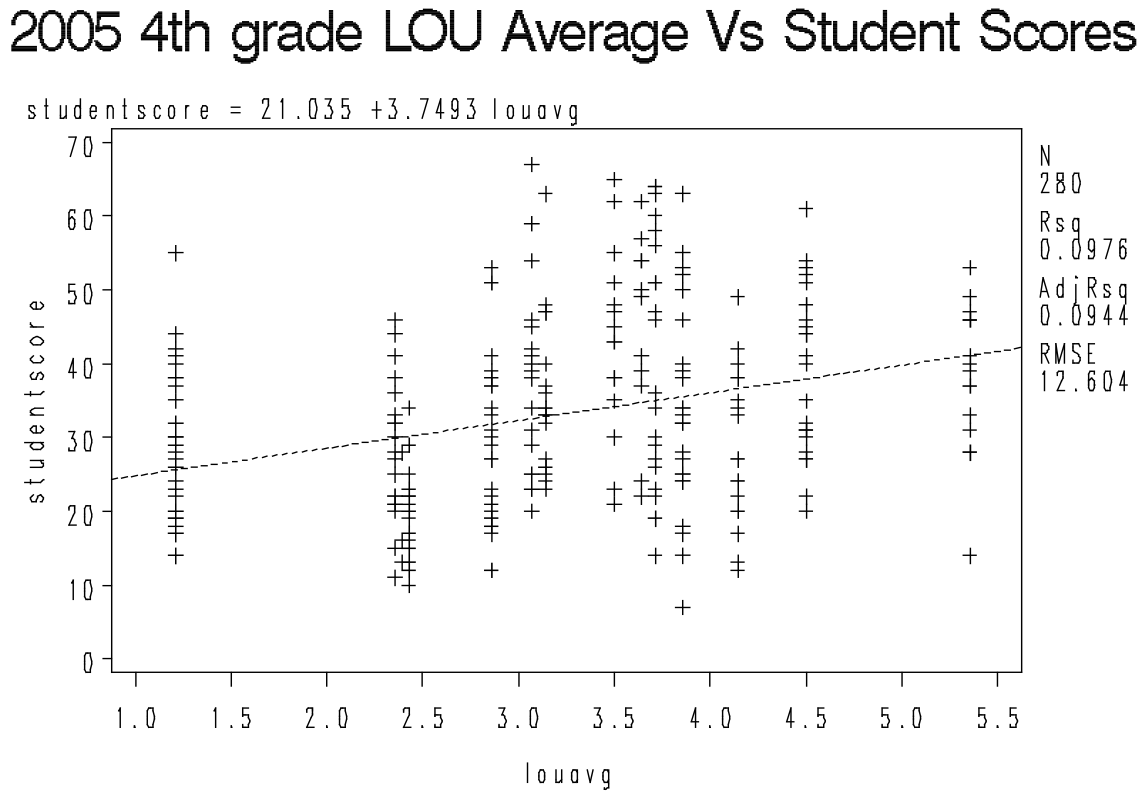


Figure 4 – Scatter plot of the students' achievement on the Criterion Reference Test of New Mexico and the level of implementation of the professional development in the classroom as measured on the Levels of Use (LOU) scale for the 4th Grade of the Gadsden Independent School. District in 2005 as part of the Gadsden Mathematics Initiative. (The LOU scale was designed specifically for the Gadsden Mathematics Initiative based on the Southwest Educational Development Labs guidelines for the design of a Concerns-based Assessment (CBAM) to measure the effect of innovation in teachers.)

No differences were detected in student achievement for 8th grade teachers in as a function of implementation (LOU) or classroom observations (CLO). This is may be due to the small number of eighth grade teachers available for observation.

<u>Research Question 1 – 2006</u>

Unfortunately, this nice linear relationship between the level of implementation of the GMI and student achievement did not continue in the 2006-2006 year. There might have been a number of reasons for this. Based on our experiences in working in the district without the strong support of the previous superintendent and associate superintendent we were aware of a decrease in implementation in some schools. When visiting schools in Spring 2006 for classroom observations we were sometimes told that little or no math at all was being taught on specific days. While some schools, like Anthony Elementary, continued to push a rigorous mathematics program and had students scoring at the 68th percentile in average mathematics scores, other schools fell behind. What is interesting is that the achievement scores as a whole did not fall. Perhaps the stronger and more effectively run schools brought up some of the less effective schools. Now that the old administration is back there has been an increase in the level of implementation of the mathematics initiative and teachers and schools now believe the initiative is here to stay. We will have to wait for the 2007 test results to see if achievement continues to improve with the returning support of upper administration.

Research Question 2. *What do we know about the changing nature of pedagogy in classrooms as a result of the GMI implementation in the 4th and 8th grades?*

2a) What do classroom observations on the Classroom Observation Snapshot tell us about types of instruction, student engagement and type of learning, and how students are learning mathematics?

2b) What relationships, if any, exist between the level of teacher use of professional development (as measured on LOU scale) and factors related to teacher classroom practice (as measured by the CLO)?

Research Question 2a – 2003-2006

Information from the Classroom Observation Snapshot on the types of instructional events in which students and teachers engaged every five minutes is provided in Table III below for all years of the study. Some ratings included more than one type of instructional activity. For example, teachers might be assessing student learning (A) while they were engaged in interacting with students (TIS).

Table III. Number of minutes and proportion of class session in which students and teachers were engaged in various instructional activities.

	Instructional Activities[1]								
Year	**L**	**LQ/CD**	**PM**	**IS**	**SP**	**SGD**	**TIS**	**A**	**INT**
2003	3	43	22	50	14	79	89	27	1
	1%	13%	7%	15%	4%	24%	27%	8%	0%
2004	15	149	76	77	47	129	154	46	3
	2%	21%	11%	11%	7%	19%	22%	7%	0%
2005	8	77	40	45	9	74	98	17	7
	2%	21%	11%	12%	2%	20%	26%	4%	2%
2006	27	162	74	106	29	89	128	15	60
	4%	23%	11%	15%	4%	13%	19%	2%	9%

[1]L = Lecturing, LQ/CD = Lecture Questions/Class Discussion, PM = Problem Modeling, IS = Independent Seat Work, SP= Student Presentations, SGD = Student Group Discussions, TIS = Teacher Interacting with Students, A=Assessing Students, and INT = Interruptions to class.

From this table it can be seen that the types of instructional strategies remained fairly consistent once the grant was implemented. The largest increase in instructional events occurred in the increases in instruction that involved *Lecture with Questions*. The Lecture with Questions was added to our observational tools during the first year of observations, as described earlier, since most teachers had learned to ask questions rather than simply lecture after the first two years of the GMI implementation. Observers noted that teachers usually began their classes with questions to students, about what they knew about the previous day's work and what they might predict about the new lesson. Students were asked to define shapes or provide answers based on their current work. The kind of instruction that was emerging in the GMI was quite different from the traditional mathematics teaching model found in many U.S. classrooms as observed on the TIMMS videos (Hiebert & Stigler,1997). The typical U.S. Mathematics Classroom method is usually something like 1) go over the homework; 2) teacher provides information through lecture about a new topic, 3) students do practice problems and are assigned homework. In contrast in countries such as Japan, the teacher might begin by presenting students with a problem and asking them to work on solving it either individually or in small groups. There was an increase in Problem Modeling (PM) in the observation of instructional events which may indicate a movement toward a new kind of instructional approach.

Research Question 2 b

Teachers' scores on the LOU scale and the ratings of teacher lessons on the CLO were highly correlated in 2003 (r = 0.91) for all years of the study. Teachers who were implementing the professional development in their classrooms at a high level were also demonstrating high-level mathematics lessons. However, the level of implementation which was measured with an instrument designed specifically to reflect the goals of the GMI seemed to be the post predictive of student success and learning especially in 2004-2005.

The classroom snapshot tool also provided very interesting information about the students' level of engagement and the type of learning processes in which they were involved during the observation. The following categories were used to describe student engagement and the types of learning in which students were involved.

<u>Student Role- Engagement</u>

HE high engagement, 80% or more of the students engaged
ME mixed engagement
LE low engagement, 80% or more of the student's off-task

<u>Learning Processes</u>

1 Receptive (lectures, observing presentations)
2 Structured Practice / Application / Review (recitation, worksheets, lecture w/questions)
3 Construction of Knowledge (organizing, describing, investigating, inventing)
.

<u>Student Engagement and Learning Processes</u>

Students were highly engaged during most of the lessons. The following data was reported for teachers in 2006 by Ken Korn, an evaluator who analyzed the observations collected in that year.

Mathematics Classroom Observation Instrument
Classroom Snapshot
Student Role

- **Based on data from the 32 Mathematics Classroom Observation Instruments I have been given, almost 2/3 of the students observed were consistently highly engaged during classroom instruction.**

(n=32 Observations)

Number and Percent of Each Kind of Student Engagement
Based on Simple Modal Analysis of Data

High Engagement	Mixed Engagement	Low Engagement
21 / 66%	8 / 25%	3 / 9%

In 2005, an analysis was done based on a review of the observation data collected by Karin Wiburg and in that years 209 of the 314 ratings on levels of engagement the observer as *High Engagement*. So 67% of the time the class was observed students were highly engaged in learning mathematics, again at least 2/3 of the time the students were highly engaged in the learning task.

The observers also paid special attention to the learning processes in which students were involved in terms of whether the learning involved receptive behaviors (listening to lectures), structured practice (practicing new concepts and procedures), and or participating in opportunities to make meaning of mathematics and construct their own understanding of the math concepts (construction). The GMI professional development was supported by the curriculum. The two textbooks chosen for use in the district, Investigations, as the elementary school, and Connected Mathematics, at the middle school assisted teachers by providing materials and guidance that helped students engage in solving mathematics problems. While, the curriculum by itself can not change learning, in combination with mathematics coaches, modeling, and good professional development it can be seen that a new type of mathematics teaching and learning was occurring in Gadsden.

2004-2006 Type of Learning Processes and Percentages of Student Time

Table 4. Percentages of student time spent in each of type of learning process by year during the Gadsden Mathematics Initiative**.**

Type of Learning Processes	**2003**	**2004**	**2005**	**2006**
1. Receptive: lectures, observing teacher presentations	**20.7%**	**4.55%**	**12.5%**	**26%**
2. Structured Practice/Applications? Review: practice, related to concepts being studied, worksheets', recitation	**38.9%**	**38.64%**	**31.3%**	**40%**
3. Construction of Knowledge: organizing, describing, investigating, inventing	**40.2%**	**56.81%**	**56.2%**	**34%**

Again there is a decrease in the amount of instructional time spent on the construction of knowledge by students in Gadsden in 2005-2006 when there was a change in administrative leaderships. In many conversations and meetings that the research team had with teachers throughout the project there was always a concern expressed about the students having enough practice with computation. Teachers may have taken advantages of a lack of oversight to ensure that students get more practice time and more drill. This could also simply have been a reflection of which teachers were observed in that year, a growth in new teachers just learning the program or changes in teacher assignments and placements in the district.

Research Question 3 *Does the type of professional development (Lesson Study or Other) or the amount of professional development affect the level of student achievement as measured on norm referenced mathematics tests administered by the state in 4th and 8th grades?*

Research Question 3 – 2003 – Does the type of professional development matter?

The numbers participating in Lesson Study were small in the first academic year of the GMI (2002-2003). The three eighth grade teachers at Santa Teresa were all doing Lesson Study. Seven other teachers in the Gadsden district were involved in Lesson Study mostly at San Miguel and Berino Elementary. Comparing the mean student test scores on the norm-referenced mathematics achievement test for the students of Lesson Study and Non-Lesson Study teachers in Gadsden, no differences were detected. However, descriptive statistics suggested a decrease in the variance in Lesson Study teacher's classrooms.

Table 5. Mean and standard deviation of student mathematics achievement scores for classrooms taught by teachers participating in Lesson Study and other (non-lesson study) types of professional development

Year	Type of PD	Student Math Achievement Score		Number of Teachers
		Mean	SD	
2003	**Non-lesson study**	38.0	14.4	25
	Lesson Study	41.6	8.8	10
2004[1]	**Non-lesson study**	363.6	32.16	36
	Lesson Study	388.4	23.71	5
2005	**Non-lesson study**	38.0	6.95	25
	Lesson Study	41.6	5.85	7
2006	**Non-lesson study**	37.3 ?	5.96	21
	Lesson Study	41.0	5.97	9

[1]In 2004, the … (Something about exam change in 2004)

In 2004, a comparison of students in classrooms taught by teachers who learned Lesson Study to those in classrooms taught by non-lesson study teachers was performed in 2004 using individual student scores (rather than classroom averages) and the hierarchical mixed linear model. In 2003-2004, 17 teachers were engaged in some form of lesson study and one school, San Miguel used a whole school lesson study model. Students in classrooms taught by teachers in the Lesson Study professional development program achieved higher scores on the state standardized test than those in classrooms where the teacher received other types of professional development. The mixed model also indicated a difference in the variability within classrooms for these two groups; Lesson Study classrooms had smaller variances than non Lesson Study classrooms.

Findings in terms of lesson study versus non-lesson study teachers continued to show significant differences throughout the rest of the study for the 4^{th} grade teachers and their students. In 2005 and in 2006, students in classrooms taught by teachers in Lesson Study PD, achieved higher math achievement scores for students in the 4^{th} grade, but this finding was not true for 8^{th} grade students in 2005 or 2006. The number of 8^{th} grade teachers involved in Lesson Study as a result of some teachers changing grades and schools was really too small in both years to measure significance.

Movement of Professional Development from Whole District to School-Based.

It should be noted that during the GMI the first two years of professional development were basically district wide professional development. In 2004-2005 and in 2005-2006 PD began to move to the school sites. In 2004-2005 each school had math mentors, although these mentors had regular classroom teaching duties they received a stipend to help coordinate the GMI efforts at this school site. In 2005-2006 full-time Mathematics Process Trainers (MPTs) were placed in the schools, one for every two elementary schools and one at each middle school. The elementary principals found these specialists to be of great assistance in implementing the mathematics program. This network of specialists received intensive training from an elementary and middle school district mathematics curriculum and strengthened the effect of the GMI in each school. However, having a mathematics specialist only half of the time was frustrating for the elementary schools and there were often conflicts over how much of the MPTs time was used at the partner schools. In 2006-2007 the decision was made to put a MPT at every K-8 school and eventually by 2007-2008 to put a math trainer in the three high schools. All of these math specialists are now supported with district operational funds ensuring that the GMI moved from being a grant-funded project to a district initiative.

Hours of Professional Development

The research literature has often reported that hours of professional development should have a significant effect on teacher quality and student achievement so we also tested this variable. Common recommendations are that teachers need at least 130 hours of professional development before there can be an effect on practice and student learning. During the summers and school years of 2002-2003 and 2003-2004, 80 hours of professional development was provided to all teachers through a commonly-structure summer academy and specific workshops during the school year. Most of the professional development focused on how to use the new standards-based curriculum materials and how to align teaching, standards, and assessment. Principals could request additional help at their school site with Professional development and asked for help on such topics as how to align instruction with testing and how to facilitate teachers working together.

During the first two years of the Student Outcomes Study, the type of professional development, specifically a type, Lesson Study or Unit Study that encouraged teacher collaboration and alignment of efforts had more effect on achievement than the hours of professional development. Follow-up conversations with GMI and Gadsden staff, who provide frequent professional development in the district for mathematics, confirmed that in those schools in which teachers worked together to learn and plan for mathematics instruction, whether using Lesson Study or another form of mathematics group study, student learning and performance are higher than in those schools where teachers are more or less working independently as math teachers.

In fact, the only variable that turned out to be a significant factor in predicting student achievement in the 2003-2004 study was whether or not teachers had participated in lesson study professional development. However, one might consider that teachers who agreed to participate in the demands of doing lesson study may simply be more dedicated teachers spending more time than other teachers on improving their own instruction. We would need to have data on how teachers and their students performed prior to the GMI and we don't have this data so caution is advised.

However, as professional development moved into the teacher's schools it had a bigger impact on student achievement. Based on data from the 2005-2006 year, for the first time total hours spent by teachers in professional development was related to student achievement for 8th

grade and 4th grade teachers and students. This could be explained by the fact that professional development during the last year of the project was done at the school site and focused more on the specific mathematics content teachers in a school were teaching. Professional development was also supported in an intensive and daily manner by the presence of a mathematics coach for every school in the district. The teachers helped develop the PD with the Mathematics Process Trainer and it was closely aligned to the concerns of the teacher at that site. This is an expensive type of professional development but if it can effectively close the achievement gap and prepare more students to continue into advanced mathematics and science classes it could be seen as cost effective for the United States.

Decreases in Variability and Achievement

It can be shown that students' scores in mathematics are continuing to increase in the Gadsden district. One way to look at general improvement in student achievement in the GMI is to consider the decrease in variance among students within the classrooms and the decrease in variance among classrooms in the district. As teachers improved their ability to engage over 80% of their students in mathematics, as indicated on classroom observations, the variability of the scores decreased. The lower students moved closer to the higher students and the result was an overall increase in mean achievement scores. Thus the districts increased achievement by bringing up the lower students and decreasing the gap between high and low students.

When analyzing district achievement data in other districts it is striking to see an increase in the gap between low and high students that grows as the students progress through the grades. Whatever happened in GISD did result in closing the achievement gap and is worth further study and efforts at replication.

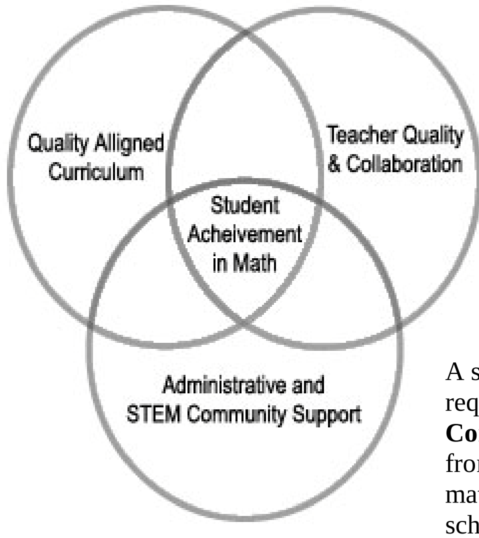
The good news for the Gadsden Mathematics Initiative is that students in Gadsden scored as well as students in Las Cruces, Albuquerque and the state in general. Something that hadn't happened in many years until the GMI was funded. The standards-based curriculum implemented in the border Gadsden School District and now heavily supported by district funds is strong. Even a problem with the school board and an upcoming recall election seem to have had no effect on the mathematics initiative.

As a result of the work in the GMI and subsequent work in other districts we are beginning to evolve a model that might be useful for building capacity for achievement. A graphical representation of the model is presented below.

A Model for Building Capacity in Student Mathematics Achievement

.

Aligned Curriculum
includes data-driven analysis of learning needs; alignment of standards, instruction, assessment and accountability from local to state level; and use of quality curriculum materials



Teacher Quality
knowledge of math content and expanded pedagogy through quality collaborative professional development

A successful mathematics program requires **Administrative and Community support** including leadership from educational researchers, mathematicians, math educators and school leaders.

Suggestions for further Research and Action

<u>School & District Effectiveness</u>. As a result of this research and our reflections on it and other new mathematics projects, there are many areas that can be suggested for further research. Differences between schools in terms of their effectiveness in implementing mathematics education reform were noticed, but we remain unsure about how these differences might be measured. Toward the end of the project, and supported by the findings related to the effect of teachers who did Lesson Study on student achievement, we developed a growing notion that the level of collaboration between teachers in a school really mattered. There is emerging research which has suggested that effectiveness of the school as well as the teacher makes a large difference in student achievement. It is really common sense that if students have access to teachers who are teaching quite differently students may not have equal access to the same rigorous mathematics learning experience. In the MC^2 project, we decided a year ago to work with only those schools and districts that are committed to fully and evenly implementing a new mathematics curriculum and instructional approach. Our districts sign memorandums of agreement for all teachers and administrators to attend training and to provide time for teachers during school to collaborate for at least 4 hours every month.

Research on the role of administrator in building teacher collaboration is needed as well as how administrators can serve as instructional leaders in mathematics education.

Mathematical Sophistication

We discovered while observing in classrooms that there were differences in teacher's knowledge and confidence in mathematics. Observers noticed that a teacher might be uncomfortable in answering a student's question when it required talking about mathematics in ways that had not been practiced. Teachers also missed opportunities to extend the mathematics in a lesson. One example occurred when students were talking about practical applications of operations. One student said in response to a cue to talk about division, "Supposed you had four pieces of pizza and six kids." The teacher answered by saying that usually in division we divide the smaller numbers into larger numbers.

Additional research is needed on teacher's background and abilities in mathematics, particularly in terms of their knowledge for teaching mathematics. We did not have adequate research personnel or time to gather data about the teachers backgrounds in mathematics. While we did add some questions related to mathematics courses and professional development experiences to our pre-interviews prior to classroom observations we found this was NOT a good time for asking teachers to reflect on mathematics. In the future we would like to do separate interviews with teachers related to their mathematics backgrounds, how they think and feel about teaching mathematics, and what they think of the classroom support for mathematics teaching provided by the district. While the questions will be similar our methodology needs to change in order to gather this information correctly.

The following questions were discussed with the district evaluator who thought they were good questions and would elicit the additional information we needed. We are going to be adding a question related to how teachers think and feel about teaching mathematics.

1. How long have you been teaching? How long have you been teaching at this school?
2. How long have you been teaching mathematics? (for 8^{th} grade teachers specifically)
3. What mathematics courses have you taken?
4. What kind of school-based professional development and/or support have you received in mathematics (lesson study, curriculum alignment and planning, instructional strategies?

5. How do you feel about teaching mathematics? What do you think are the most important things to consider when teaching mathematics?

English Language Learners and Success in Mathematics

Gadsden had remarkable success in helping ELLs to success in mathematics. Teachers in the district have developed interesting methods for engaging their students in mathematics and some cases for using the students' native language, Spanish, to help students work at a higher level conceptually in mathematics than might be expected by teachers who do not understand the nature of language and learning. Rocio Benedicto, a doctoral research assistant who worked on the GMI, will be completing a dissertation during the next year related to how the TESOL and Bilingual teachers helped ELL students to be successful in mathematics. Research has shown in other cases that using NSF-approved curricula like *Investigations* and *Connected Mathematics* has resulted in additional success for diverse students.

Other districts in New Mexico, the Public Education Department and national groups have become interested in how this success in mathematics was achieved in what has been called the "Gadsden Model". The work continues in this district to refine the model and continue to optimize mathematics learning for all students.

References

Glaser, E. (2005) The Development of a Mixed-Effects Model for Evaluating the Effectiveness of the Gadsden Mathematics Initiative, Unpublished Masters Thesis

Hill, H.C.; Rowan,B; Ball, D.L. (2005) Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement, American Educational Research Journal, v42 n2 p371-406 Sum 2005

Kramarski, B., Merarch, Z.R. (2003). Enhancing Mathematical Reasoning in the Classroom: The Effects of Cooperative Learning and Metacognitive Training. *American Educational Research Journal*, vol., 40, no. 1, 281-310.

Maxwell, A. J. (2004). Causal Explanation, Qualitative Research, and Scientific Inquiry in Education. *Educational Researcher*, vol. 33, no. 2, 3-11.

Hiebert, J. & Stigler, J.W. (1997) Understanding and Improving Classroom Mathematics Instruction: An Overview of the TIMSS Video Study, Phi Delta Kappan, Vol. 78

Wiburg, K & Brown, S. (2007) Lesson Study Communities: Increasing Achievement with Diverse Students. Corwin Press.

Appendix A – Instruments Used in the GMI Student Outcomes Study

<u>I. - Classroom Snapshot</u> – The trained observer used the following classroom snapshot to indicate the type of instruction, learning , and tools used in the classroom by writing down every 5 minutes exactly what was going on. A new category was added to this scale called Lecture with Question since students were doing most of the answering by the teacher had structured the questions in a lecture-like way.

Classroom Checklist Instructions:

Please fill in the instructional strategies (<u>not</u> the instructor's actual activities, in case they are correcting papers or something non instructional), student role, and learning processes used in each five-minute portion of this class in the boxes below. There may be one or more strategies used in each category during each interval. For example, SGD, TIS, and A often occur together in a five-minute period, but SGD and L do not.

<u>Type of Instruction</u>

L lecture/presentation
LQ lecture w/questions
PM problem modeling/demonstration
CD class discussion
IS independent seatwork
SP student presentation
SGD small group work/discussion
TIS teacher/instructor interacting w/student
A assessment
I interruption
AD administrative tasks

<u>Student Role</u>

HE high engagement, 80% or more of the students engaged
ME mixed engagement
LE low engagement, 80% or more of the students off-task

<u>Learning Processes</u>

1 Receptive (lectures, observing presentations)
2 Structured Practice / Application / Review (recitation, worksheets, lecture w/questions)
3 Construction of Knowledge (organizing, describing, investigating, inventing)

<u>Use of Tools</u>

C: Calculator CS: Computer/Software / Application I: Internet / Web Resources
OL: Online Communication V: Video M: Manipulatives O: Other____________

A: Available, but not used
SB: Used for skill building (drill and practice)
PS: Used as a problem solving tool in developing content knowledge
(Use two codes in this box, one for the type of technology use and other for type of application.)

Time in Minutes:

	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
Instruction												
Student												
Learning												
Tools												

Modified from CETP – Core Evaluation Classroom Observation Protocol which was adopted from Horizon?

II- The Levels of Use Scale development and Instrument used

II. The Levels of Use (LOU) Scale was based on the Guide from the Southwest Educational Development Guide for measuring the integration of an innovation. Karin Wiburg and Wanda Guzman (Temaz) looked carefully at the intended outcome of the GMI professional development as well as the levels of use used in the original Horizon instrument for measuring the integration of an innovation in the Local Systemic Change initiatives. We used the Levels from 0 to 6 with 0 indicating almost no effect of the professional development in the classroom and 6 indicting that the teaching and learning in the classroom is exactly what was intended as a result of the professional development provided. The Categories (0-6) and the Descriptions of observable behavior related to each category rating is described on this and the next page.

Level	Category	Description of Behavior Observed
0	Nonuse Level 0	Little or no evidence of student thinking or engagement with important ideas of mathematics. Instruction is characterized as one of the following: • Passive “Learning: Students are passive recipients of information from the teachers or textbook; material is presented in a way that is inaccessible to many of the students. • Activity for Activity’s Sake: Students are involved in hands-on activities or other individual or group work, but the lesson lacks a clear sense of purpose and/or a clear link to conceptual development of mathematics. Overall, the lesson is highly unlikely to enhance students’ understanding of mathematics or develop the capacity to successfully “do” mathematics.
1	Awareness Level 1	Instruction contains some elements of effective practice, but there are serious problems in the design, implementation, content, and/or appropriateness for many students in the class. The lesson design may include elements of standards based instruction, but instruction is primarily teacher centered and emphasizes learning procedures for solving computation problems rather than an exploration of mathematics concepts. Overall, the lesson is very limited in its likelihood to enhance students’ understanding of mathematics or develop the capacity to successfully “do” mathematics.
2	Exploration Level 2	Elements of the lesson reflect student-thinking/student-centered instruction and the teacher makes use of investigative curriculum resources. However, instruction is still characterized as the teacher of the primary provider of knowledge. Students are, at times, engaged in meaningful work, but there are substantial weaknesses in the design, implementation, and/or content of the instruction. For example, during an exploration the teachers may tell students the procedure they should be using to solve the problem. Teacher does not assess effectiveness of lesson in terms of developing student understanding. Instruction may not adequately address the needs of a number of students. Overall, the lesson is limited in its likelihood to enhance students’ understanding of mathematics or develop the capacity to successfully “do” mathematics.

Number	Level	Description of Behavior Observed
3	Infusion Level 3	The lesson reflects student-thinking/student-centered instruction and makes use of investigative curriculum resources. Students are, at times, engaged in meaningful work, but there are some weaknesses in the in the design, implementation, and/or content of the instruction. For example, the teacher may short-circuit a planned exploration by telling students what they "should have found". Teacher does not fully assess effectiveness of lesson in terms of developing student understanding, but is more concerned with student engagement. Implementation of curriculum resources is mechanical, and may not adequately address the needs of all students in the class. Overall, the lesson is somewhat limited in its likelihood to enhance students' understanding of mathematics or develop the capacity to successfully "do" mathematics.
4	Integration Level 4	The lesson design and implementation incorporates a student-thinking/student-centered model and makes use of investigative curriculum resources. Students are engaged in meaningful work most of the time, but there are minor weaknesses in the in the design, implementation, and/or content of the instruction. Implementation of curriculum resources is still somewhat mechanical, but the teacher is able to reflect on the lesson's effectiveness in terms of developing student understanding. Overall, the lesson is limited in its likelihood to enhance students' understanding of mathematics or develop the capacity to successfully "do" mathematics.
5	Expansion Level 5	Instruction is purposeful and engaging for most of the students. Students actively participate in meaningful work that builds conceptual understanding of mathematics. The lesson is well designed and the teacher implements it well, but adaptation of content or pedagogy in response to students' needs or interests is limited. Teacher assesses the effectiveness of lessons by assessing student understanding of concepts. Instruction is quite likely to enhance most students' understanding of mathematics and develop the capacity to successfully "do" mathematics.
6	Refinement Level 6	Instruction is purposeful and all students are highly engaged most or all of the time in meaningful work that builds conceptual understanding of mathematics. The lesson is well designed and artfully implemented, with flexibility and responsiveness to students' needs or interests. Assessment and instruction are fully integrated and the teacher assesses the effectiveness of lessons by assessing student understanding of concepts. The teacher is able to independently develop lessons that are centered on a mathematical problem. Instruction is highly likely to enhance most students' understanding of mathematics and develop the capacity to successfully "do" mathematics.

II - Levels of Use Ratings of Key Indicators (LOU)

In this section, you are asked to rate each of a number of key indicators as descriptive of the lesson in five categories, from 0 (not at all) to 6 (to a great extent). Note that any one lesson is not likely to provide evidence for every single indicator; use DK, "Don't Know," when there is not enough evidence for you to make a judgment. Use N/A, "Not Applicable," when you consider the indicator inappropriate given the purpose and context of the lesson.

1. The lesson design promoted strongly coherent conceptual understanding of mathematics	0 1 2 3 4 5 6 DK N/A
2. The teacher's instructional decisions and questions enhanced student's conceptual understanding of the mathematics	0 1 2 3 4 5 6 DK N/A
3. Elements of abstraction (i.e. symbolic representations, theory building) were encouraged when it was important to do so	0 1 2 3 4 5 6 DK N/A
4. Students were reflective about learning	0 1 2 3 4 5 6 DK N/A
5. The lesson was designed to engage students as members of a learning community	0 1 2 3 4 5 6 DK N/A
6. The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein	0 1 2 3 4 5 6 DK N/A
7. Interactions reflected collaborative working relationships among students (e.g. , students worked together, talked with each other about the lesson), and between teacher/instructor and students	0 1 2 3 4 5 6 DK N/A
8. Intellectual rigor, constructive criticism, and the challenging of ideas was valued	0 1 2 3 4 5 6 DK N/A
9. Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence	0 1 2 3 4 5 6 DK N/A
10. The teacher/instructor displayed an understanding of mathematics concepts (e.g., in his/her dialogue with students)	0 1 2 3 4 5 6 DK N/A
11. Appropriate connections were made to other areas of mathematics, to other disciplines, and/or to real-world contexts, social issues, and global concerns	0 1 2 3 4 5 6 DK N/A

For the following questions, select the response that best describes your overall assessment of the *likely effect* of this lesson in each of the following areas.

12. Students' understanding of mathematics enhanced and enriched by investigation	0 1 2 3 4 5 6 DK N/A
13. Students' understanding of important mathematics concepts	0 1 2 3 4 5 6 DK N/A
14. Students' capacity to carry out their own mathematical investigations	0 1 2 3 4 5 6 DK N/A

Modified from CETP – Core Evaluation Classroom Observation Protocol

III- <u>Classroom Lesson Observation Worksheet (CLO)-</u> this form corresponds to the lesson observation protocol used by the Horizon Evaluation Tool. Trainers needed to rate lessons approximately the same as experts in math or science teaching after watching videos and then real lessons. Use of this tool for evaluation required the most training. Lack of funding for research in the grant caused us to train graduate students ourselves and this may be the most unreliable measure in the group of measures.

Mathematical Focus of the Lesson:	
Lesson Design: 0 1 2 3 4 5 6	*Strengths:* *Weaknesses:*
Lesson Implementation: 0 1 2 3 4 5 6	*Strengths:* *Weaknesses:*
Mathematics Content: 0 1 2 3 4 5 6	*Strengths:* *Weaknesses:*
Materials and Resources: 0 1 2 3 4 5 6	*Strengths:* *Weaknesses:*
Classroom: Culture/Environment 0 1 2 3 4 5 6	*Strengths:* *Weaknesses:*
Overall Rating: 0 1 2 3 4 5 6	*Strengths:* *Weaknesses:*

Modified from CETP – Core Evaluation Classroom Observation Protocol

Appendix B - Development of the Mixed Effects Model for the GMI

A review of the educational research literature for academic year 2003-04 showed that regression, correlation analysis, and analysis of variance are commonly used methods for statistical analysis. In dealing with student achievement structural equation models are often used. From a statistician's perspective, the dominant feature of student achievement data is its hierarchical structure. Students are nested within classrooms, which in turn are nested within a school[1]. However, mixed-effects models seem to be just beginning to emerge into use. A later review of the literature confirmed that mixed-effects models are becoming a more common analytic method in education research.

The term *mixed-effects model* should not be confused with the education research term *mixed methodology*, referring to a study that uses both quantitative and qualitative methods. A mixed-effects model is a linear statistical model that includes both fixed effects and random effects as experimental factors. If a study factor includes all possible levels of the factor (or at least all the levels which we intend to make inferences on) it is called a fixed effect. If the study factor includes a random sample of all possible levels it is called a random effect. (Littell, et al. 1996)

This mixed-effects model is a member of the more general class of linear hierarchical models. These models are seeing increasing use in the analysis of educational data. They arise naturally in this context as the result of the sampling mechanism. Selecting students at random from among a broad class (say, all New Mexico fourth grade students) would be difficult at best. Assigning these students to receive independently assigned educational factors would be a nightmare. Instead, the students are best served by sampling from the existing hierarchy of State, districts, schools, and classrooms. Sampling using the hierarchical structure requires analyses that take the sampling schema into account.

The specific mixed-effects model used in the GMI was developed by Erica Glaser under the direction of Professor Marta Remmenga, a member of the University Statistics Center at New Mexico State University. The details of her development are available by request from Dr. Remmenga. The model developed treats teachers as a random factor nested within the school, and models student achievement scores as a function of either lesson-study or level-of-use (fixed effects) and random effects due to teacher (nested within school) and the student herself (random variation in the population of students). Glaser performed a fairly extensive model residual analysis to assure that the model is allowing for natural features of the data. For example, the residual analysis revealed that the classroom variances appear to be smaller in the lesson-plan group than in the other-development group. She modified the model to allow for the heterogeneous variances. Variation among the teachers makes a contribution to the variability of the estimated means (for lesson-study) and slope (for level-of-use). All analyses are carried out using the SAS® MIXED procedure (SAS Institute, 2004).

[1] Conceptually, this nesting continues in larger studies as schools occur within districts; districts are then nested within a State.