MATHEMATICS TEACHER TRANSFORMATION INSTITUTES (MTTI)
Lehman College, The City University of New York
Highlights from Year 1 Research Activities

OVERVIEW
Lehman College's **Mathematics Teacher Transformation Institutes (*MTTI*)** program is designed to support the development of teacher leaders in mathematics in Bronx middle and high schools. Two cohorts, each comprised of 40 experienced Bronx middle and high school teachers, will be engaged in three mutually-reinforcing components for a period of three years—coursework in advanced concepts in algebra and geometry; classroom-based inquiry; and formal leadership training along with consulting on-site at participating teachers' schools. As a multi-faceted collaboration among College faculty in mathematics and mathematics education, staff-development specialists from the New York City Mathematics Project, a program of Lehman's Institute for Literacy Studies, and NYC Department of Education leaders, MTTI's primary goal is to enhance student achievement in mathematics by improving teaching and learning at school and college levels.

HIGHLIGHTS
These highlights present baseline data collected during the first year of this five-year project, October 1, 2008 through September 30, 2009. The data offer information about Cohort 1—Bronx middle and high school teachers—general characteristics of the schools and students they serve, and some preliminary findings from measures we are using to assess:
- the relationship over time between teacher knowledge, pedagogical change, and student performance in mathematics;
- factors that influence the development of teacher leaders over time;
- factors that influence change in curriculum in diverse institutional contexts over time.

Since we are conducting a longitudinal study, we are documenting and tracking the evolution of teacher leadership from three vantage points:
- how teachers engage their students in conceptually-challenging mathematics;
- how Learning Trajectories from classroom-based inquiry help teachers to identify and address pedagogical issues;
- the extent to which formal elements of leadership training contribute to change in diverse institutional contexts.

By L*earning Trajectories*, we mean the conjectures that teachers make about possible learning routes that aim at significant mathematical ideas and a specific means that can be used to support and organize learning along this route. We assume teachers design instructional tasks with guidelines that suggest an order for the tasks and expectations

about the types of thinking and learning in which students engage as they participate in the instructional tasks *(The Teaching Trajectory)*.

As of January 2009, MTTI recruited a total of 42 participants for Cohort 1—representing 32 Bronx schools. To participate in this program, participants must have at least four years of teaching experience, permanent certification as a mathematics teacher, and a recommendation from the principal of their school. Table 1 presents relevant descriptive data of the participating teachers who began the MTTI program in January 2009. Data describing the participating public schools appear in Table 2. Table 3 presents the performance of students from these schools on standardized tests.

Table 1. Cohort 1 Participants as of January 2009

Bronx Middle School Teachers	13
Bronx High School Teachers	29
Average # of Years Teaching	8.5
# Teachers, Undergraduate Math Major	20
# Teachers, Double Major (including Math)	3
# Teachers, Major with significant math content	9
# Teachers, Graduate Training in Education	35
# Teachers, Graduate Training in Math Ed	15

Table 2. Cohort 1 Schools as of January 2009: Descriptive Statistics

Descriptor	Middle Schools	High Schools
Number of Schools	12	20
Enrollment	545	405
Math Class Size	28	25
Percent Eligible for Free Lunch	57%	69%
Percent Eligible for Reduced Price Lunch	9%	7%
Percent Student Stability	83%	74%
ELL Students	13%	11%
Black	35%	36%
Latino	46%	59%
White	6%	2%
Asian	6%	2%

Table 3. Baseline Student Performance on New York State Assessments: Percent of Students Passing Regents Math A or Math B at Three Levels of Performance

	Middle Schools	High Schools
Math A 2007, Percent 55	91%	74%
Math A 2007, Percent 65	88%	47%
Math A 2007, Percent 85	51%	4%
Math B 2007, Percent 55	na*	56%
Math B 2007, Percent 65	na	44%
Math B 2007, Percent 85	na	8%

*[Middle school students do not take Math B.]

METHODS

Because we are addressing a wide range of questions, a number of statistical approaches are being employed to analyze the various data sets we have collected to assess teachers' progress over time. At this point, only baseline data have been collected, the highlights of which are being presented in the remainder of this document.

The overarching questions we are investigating and the relevant data sources that we are using to analyze the data for each question are as follows, with questions underlined, followed by the data sources.

1) The relationship among teacher knowledge, pedagogical change, and student performance in mathematics over time. Teacher-knowledge measures; classroom observations; lesson plans; teachers interviews; worksheets; teacher surveys; student standardized test scores; classroom-based inquiry projects; NYCMP staff-development specialist logs.

2) Factors that influence the development of teacher leadership over time. Teacher-knowledge measures; classroom observations; lesson plans; worksheets; teacher surveys; student standardized test scores; classroom-based inquiry projects; NYCMP staff-development specialist logs.

3) Factors that influence change in curriculum in diverse institutional contexts over time. Interviews of mathematics and mathematics education faculty; interviews with teachers and school leaders; surveys of mathematics and mathematics education faculty; mathematics faculty observations; review of course materials; NYCMP staff-development specialists logs.

PRELIMINARY FINDINGS

During year one, October 1, 2008 through September 30, 2009, we concentrated on establishing baseline data for Cohort 1 to consider the relationship among teacher knowledge, pedagogical change, and student performance in mathematics over time.

Mathematics Content Knowledge

Louisville Algebra and Geometry Pre-Tests. The Louisville Algebra and Geometry (Middle School) pre-tests were administered at Lehman College to all entering MTTI participants (n=43) during January 2009, in each of two math classes. Twenty-two participants were enrolled in the Fundamentals of Mathematics course (MAT670); and, 21 participants were enrolled in the Geometry (MAT631) course. Unique identification numbers for each participant were recorded on the test papers. The pre-tests were scored by independent assessors at the University of Louisville. The Algebra and Geometry tests were taken in one of two orders—either Algebra followed by Geometry, or Geometry followed by Algebra. The order of test taking was randomized across participants.

Twenty-two participants took the Algebra test followed by the Geometry test. Twenty-one participants took the Geometry test followed by the Algebra test. Participants were told that each test should take approximately 45 to 60 minutes to complete, but they were free to allocate their time as they saw fit with the provision that the tests had to be taken in the order in which they were presented. Participants recorded their start and finish times for both tests on the test paper. After test completion, an examination of the start and finish times on the tests showed that all participants completed the tests in the order in which they were presented.

Results. The table below shows teachers' scores on both the Louisville Algebra and Geometry tests. Out of a maximum of 40 points for each assessment, the Algebra mean score was 25.1 (std. = 6.72) with a median score of 27, and the Geometry mean score was 21.6 (std. = 7.25) with a median score of 22.0. Algebra scores were significantly higher than Geometry scores.

Table 4. Algebra vs. Geometry

	Mean	Std. Deviation	Mean Difference	t
Algebra (n = 43)	25.1	6.719		
Geometry (n = 43)	21.6	7.251	3.47	4.289*

*Difference is significant (p<.01)

Comparisons of the content-knowledge items on the Louisville test (Algebra-40 items; Geometry-40 items) and for Algebra and Geometry combined (80 items) were made between the following groups (.05 level).

Math Majors vs. non-Majors. No significant mean differences were observed.

Table 4a. Math Majors (n = 25) vs. Non-Math Majors (n = 18)

	Mean Difference	t	p	Effect Size (Cohen's d)
All 40 items of the Louisville Geometry Test	.81	.357	.723	.110
All 40 items of the Louisville Algebra Test	3.04	1.486	.145	.460
z-score: All 80 items of the combined (Geometry + Algebra) Louisville tests	.2822	.985	.330	.30000

Math Education Degree vs. Non-Math Education Degree. Significant mean differences were observed in favor of teachers with a degree in Math Education on all 80 items of the combined (Algebra and Geometry) Louisville tests, and on all 40 items of the Algebra test. No significant mean differences were observed on all 40 items of the Geometry test.

4

Table 4b. Math Ed Degree (n = 23) vs. Non-Math Ed (n = 20)

	Mean Difference	t	p	Effect Size (Cohen's d)
All 40 items of the Louisville Geometry Test	4.17	1.943	.059	.590
All 40 items of the Louisville Algebra Test	5.32	2.792	.008**	.850
Zscore: All 80 items of the combined (Geometry + Algebra) Louisville tests	.6837	2.571	.014*	.78000

*Significance at the .05 level   ** Significance at the .01 level

Middle (n =13) vs. High School (n = 30). No significant differences were observed.

Table 4c. Middle (n =13) vs. High School (n = 30)

	Mean Difference	t	p	Effect Size (Cohen's d)
All 40 items of the Louisville Geometry Test	-.72	-.297	.768	-.100
All 40 items of the Louisville Algebra Test	-3.71	-1.698	.097	-.560
Zscore: All 80 items of the combined (Geometry + Algebra) Louisville tests	-.3256	-1.060	.295	-.35000

Gender Female (n = 27) vs. Males (n = 16). No significant differences were observed.

Table 4d. Gender Female (n = 27) vs. Males (n = 16)

	Mean Difference	t	p	Effect Size (Cohen's d)
All 40 items of the Louisville Geometry Test	-1.76	-.766	.448	-.240
All 40 items of the Louisville Algebra Test	-.82	-.384	.703	-.120
Zscore: All 80 items of the combined (Geometry + Algebra) Louisville tests	-.1826	-.621	.538	-.20000

Self-Report Barriers to Teaching? Yes (n = 26) vs. No (n = 17). No significant differences were observed between teachers who reported that they experienced barriers to their teaching and those who did not.

5

Table 4e. Barriers to Teaching? Yes (n = 26) vs. No (n = 17)

	Mean Difference	t	p	Effect Size (Cohen's d)
All 40 items of the Louisville Geometry Test	1.06	.464	.645	.140
All 40 items of the Louisville Algebra Test	3.77	1.852	.071	.580
Zscore: All 80 items of the combined (Geometry + Algebra) Louisville tests	.3538	1.233	.225	.39000

Order Results. Using the Algebra pre-test or the Geometry pre-test as a dependent variable, no significant differences in results were found between the order in which participants took the tests.

Table 4f. Louisville Pre-test Means by Order in which Teachers took the exam (Maximum Score = 40 points)

	N	Mean	Std. Deviation	Mean Difference	t	p
Algebra Exam then Geometry Exam	22	26.09	5.537			
				2.14	1.041	.302*
Geometry Exam then Algebra Exam	21	23.95	7.755			
Teachers Who Took Geometry Course First	21	27.05	3.866			
				3.91	1.973	.055*
Teachers Who Took Algebra Course First	22	23.14	8.259			

*Difference not significant at the .05 level. The order in which teachers took the test did not matter.

Time Needed to Complete Louisville Exams. On average, participants took 50 minutes (Max=85; Min=28) to complete the Louisville Algebra test. On average, they took 39 minutes (Max=60; Min=21) to complete the Louisville Geometry test. or about a minute per question. The time it took participants to finish the test was not correlated with their performance on the Louisville exams.

Table 4g. Louisville Algebra Time taken to complete in minutes

	n	Mean	Std. Deviation	Mean Difference	t	p
Math Ed Majors	23	49.09	12.442			
Non- Math Ed Majors	20	50.95	14.529	-1.86	.-453	-.653
Female	27	51.89	10.729	5.20	1.246	.220

	n	Mean	Std. Deviation		
Male	16	46.69	16.712		

4h. Louisville Geometry Time taken to complete in minutes

	n	Mean	Std. Deviation	Mean Difference	t	p
Math Ed Majors	22	36.50	8.022			
Non- Math Ed Majors	19	40.84	11.711	-4.34	-1.401	-.169 *
Female	26	39.15	9.350			
Male	15	37.40	11.331	.535	.596	1.75

Faculty-Designed Algebra and Geometry Pre-Tests. The mathematics faculty designed two purely content exams for Algebra and Geometry. Each exam was administered at the beginning of the program and scored two times by two different mathematics faculty with a very high correlation between the scores ($r = .951$).

Results. Comparisons at the .05 level between groups showed that no significant mean differences existed between: Math Majors and non-Majors; Math Education degree and non-Math Education degree; middle and high school; male and female; and, those who see barriers to teachers and those who do not ($p > .05$ for each t-test).

5a. Math Major vs. Non-math Majors

		Mean	Std. Deviation	Mean Difference	t
Faculty-designed **Geometry** Total Score	Math Major (n = 23)	28.17	18.312	5.73	.989*
	Non-Math Major (n = 18)	22.44	18.468		
Faculty-designed **Algebra** Total Score	Math Major (12)	36.75	18.548	3.45	.450*
	Non-Math Major (10)	33.30	17.049		

- Not significant (p > .05)

5b. Math Ed Degree vs. Non-Math Ed Degree

		Mean	Std. Deviation	Mean Difference	t
Faculty-designed **Geometry** Total Score	Math Ed Degree (n = 22)	30.41	20.442	10.25	-1.833 *
	No Math Ed Degree (n = 19)	20.16	14.261		
Faculty-designed **Algebra** Total Score	Male (n = 9)	40.75	16.521	12.25	1.703*
	Female (n = 13)	28.50	17.142		

- Not significant (p > .05).

5c. Middle School vs. High School

		Mean	Std. Deviation	Mean Difference	t
Faculty-designed **Geometry** Total Score	Middle School (n = 12)	19.00	15.057		-1.517 *
	High School (n = 29)	28.41	19.143	-9.41	
Faculty-designed **Algebra** Total Score	Middle School (n = 13)	32.69	18.490	-6.09	-.793*
	High School (n = 9)	38.78	16.453		

* Not significant (p > .05)

5d. Male vs. female

		Mean	Std. Deviation	Mean Difference	t
Faculty-designed **Geometry** Total Score	Male (n = 15)	25.20	19.940	-.72	-.120*
	Female (n = 26)	25.92	17.817		
Faculty-designed **Algebra** Total Score	Male (n = 6)	39.50	22.528	5.94	.698*
	Female (n = 16)	33.56	15.862		

* Not significant (p > .05)

5f. Barriers to Teaching (Yes vs. No)

		Mean Difference	Std. Deviation	t
Faculty-designed Geometry Total Score	Barriers to Teaching (Yes, n = 12)	-2.00	24.44	--.337*
	Barriers to Teaching (No, n = 29)		26.44	
Faculty-designed Algebra Total Score	Barriers to Teaching (Yes, n =9)	-3.32	33.22	--.427*
	Barriers to Teaching (No, n = 13)		36.54	

* Not significant (p > .05)

Louisville vs. Faculty-Designed Tests. High correlations were observed between the Louisville Algebra and faculty-designed Algebra tests (r = .702, p < .01) and between the Louisville Geometry and faculty-designed Geometry tests (r = .608, p < .01). A paired-sample t-test revealed, however, that the mean difference between the Louisville Algebra and faculty-designed Algebra was significant (p<.01), whereas the mean difference between the Louisville geometry and faculty-designed geometry was not (p>.05).

Table 6: Louisville vs. Faculty-Designed Tests

	Mean	Std. Deviation	n	Mean Difference	t
Louisville Geometry Pre Sub-categories Total/40	21.98	6.876	41		
Faculty-designed geometry Total Score/90	25.66	18.376	41	3.68	1.502
Louisville Algebra total/40	23.14	8.259	22		
Faculty-Designed Algebra Total Score/90	31.045	13.5593	22	7.909	3.571**

** Correlation is significant at the 0.01 level.

Course Grades. The spring 2009 course grades are reported in the table below. Teachers' grades significantly correlated with the Louisville Geometry (r = .571, p = .009) and the faculty-designed Geometry test (r = .458, p = .003). The correlation between the course grades and the two algebra tests was not, however, significant (Louisville Algebra r = .405, p = .068; faculty-designed r = .405, p = .068).

Table7: Course Grades Spring and Fall 2009

	Frequency	Percent
A	16	37.2
A-	10	23.3
B+	1	2.3
B	4	9.3
B-	3	7.0
C+	2	4.7
C	1	2.3
INC	4	9.3
MISSING	2	4.7
Total	43	100.0

	Faculty- Designed Geometry Total Score	Faculty- Designed Algebra Total Score	Louisville Algebra Sub categories Total/40	Louisville Geometry Pre Sub-categories Total/40
Course Grade Spring 2009	r = 571(**) p = .009 n = 41	r = .405 p = .068 n = 21	r = .244 p = .124 n = 41	r = .458(**) p = .003 n = 41

** Correlation is significant at the 0.01 level.

Pedagogical Content Knowledge

For both the Algebra and Geometry tests, one section comprising 10 items each focused on pedagogical content knowledge. Comparisons of the Pedagogical Content Knowledge items of the Louisville test (Algebra-10 items; Geometry -10 items) and the combined Algebra and Geometry items (a total of 20 items) were made between the following groups. As before, the summary statements are preceded by the tabular presentation of the relevant data.

Math Majors vs. non-Majors. No significant mean differences were observed.

Table 9a. Math Majors (n = 25) vs. Non-Math Majors (n = 18)

	Mean Difference	t	p	Effect Size (Cohen's d)
The subset of Geometry Pedagogical Content (10 items)	.20	.278	.782	.090
The subset of Algebra Pedagogical Content (10 items)	.32	.553	.583	.170
z-score: Combined Pedagogical Content (10 geometry items and 10 algebra items)	.1297	.495	.624	.15000

Math Education Degree vs. Non-Math Education Degree. Significant mean differences were observed in favor of participants with a Math Education degree on the combined Pedagogical Content Knowledge items and on the subset of Algebra Pedagogical Content Knowledge (10 items). No significant mean differences were observed on the subset of Geometry Pedagogical Content Knowledge (10 items).

Table 9b. T-Test - Math Ed Degree (n = 23) vs. Non-Math Ed (n = 20)

	Mean Difference	t	p	Effect Size (Cohen's d)
The subset of Geometry Pedagogical Content (10 items)	.48	.697	.490	.210
The subset of Algebra Pedagogical Content (10 items)	1.66	3.233	.002**	.990
Zscore: Combined Pedagogical Content (10 geometry items and 10 algebra items)	.5538	2.257	.029*	.70000

 *Significance at the .05 level   ** Significance at the .01 level

Middle (n =13) vs. High School (n = 30). No significant differences were observed at the .05 level.

Table 9c. Middle (n =13) vs. High School (n = 30)

	Mean Difference	t	p	Effect Size (Cohen's d)
The subset of Geometry Pedagogical Content (10 items)	.41	.548	.587	.190
The subset of Algebra Pedagogical Content (10 items)	-.05	-.074	.941	-.030
Zscore: Combined Pedagogical Content (10 geometry items and 10 algebra items)	.0793	.281	.780	.09000

Gender Female (n = 27) vs. Males (n = 16). No significant differences were observed at the .05 level.

Table9d. Gender Female (n = 27) vs. Males (n = 16)

	Mean Difference	t	p	Effect Size (Cohen's d)
The subset of Geometry Pedagogical Content (10 items)	.39	.547	.588	.170
The subset of Algebra Pedagogical Content (10 items)	.00	-.004	.997	.000
Zscore: Combined Pedagogical Content (10 geometry items and 10 algebra items)	.0863	.322	.749	.10000

Self-Report Barriers to Teaching? Yes (n = 26) vs. No (n = 17). No significant differences were observed between teachers who reported that they experienced barriers to their teaching and those who did not.

Table 9e. Barriers to Teaching? Yes (n = 26) vs. No (n = 17)

	Mean Difference	t	p	Effect Size (Cohen's d)
The subset of Geometry Pedagogical Content (10 items)	.63	.897	.375	.280
The subset of Algebra Pedagogical Content (10 items)	.79	1.388	.173	.430
Zscore: Combined Pedagogical Content (10 geometry items and 10 algebra items)	.3543	1.365	.180	.43000

Classroom Observations

A Classroom Observation Protocol (COP) was constructed through the selection of items that had been shown to be predictive of standards-based instruction and positive

11

student outcomes. Sources used included: Horizon Research, Inc.; the Arizona Collaborative for Excellence in Teacher Preparation; Evaluation of the Long-Term Effect of Teacher Enhancement project; the Constructing Physics Understanding Evaluation project; and the Systemic Initiatives Evaluation project. Preliminary estimates of the instruments' inter-rater and intra-rater reliabilities range from 50% to 80%, while internal consistency analyses resulted in alphas of .90 or better (Lawrenz, Huffman, & Appledoorn, 2002). The COP described and required ratings that included:

- Instructional Activities (lecture…);
- Levels of Student Engagement (high, medium, or low);
- Levels of Student Cognitive Activity (receipt of knowledge…).

In addition, the protocol also had observers:
- rate key indicators of student-centered instruction;
- comment on the effectiveness of the lesson in several categories;
- designate an overall "Capsule Rating" of the lesson's effectiveness, ranging from 1 (Ineffective Instruction) to 7 (Exemplary Instruction).

Participants were observed as they taught a mathematics lesson a minimum of three times by three seasoned former mathematics teachers who retired as Assistant Principals or Principals. In order to promote inter-rater reliability, the research team conducted two training sessions for the group of three observers, which were held at one middle school and one high school. In each school, trainees observed two classrooms and then compared their ratings. To develop a shared understanding of the ratings, differences were discussed and resolved.

> **Classroom Observation Results**. T he classroom observations conducted during fall 2009 ranged in length from 40 minutes to one hour, with the average length being 50 minutes, or ten five-minute segments. In total, 903 five-minute segments were coded and analyzed, showing 20 different types of instructional activities recorded for a total of 2,562 times. Results show that activities deemed typically traditional **(**lecture, lecture with discussion, writing work, etc**…)** were observed 1,002 times during the 902 five-minute segments, or 39% of the time. Activities that were standards-based and student-centered were observed 879 times, or 34% of the time. Activities that were partially traditional and partially student-centered were observed 608 times or 24% of the time.
>
> In contrast, the spring 2009 pilot observations of 20 teachers revealed more teacher-centered results. The lessons observed in the spring 2009 ranged in length from thirty minutes to one hour, with the average length being 50 minutes, or ten five-minute segments for a total of 167 five-minute segments that showed 17 different types of instructional activities recorded for a total of 132 times. Results showed that activities deemed typically traditional were observed

in 72 times during the 167 five-minute segments, or 55% of the time, leading us to believe, at the time, that participants' teaching styles were mostly teacher-centered. The drop in activities deemed typically traditional could be due to the work of MTTI consultants with participating teachers, which began in the spring but was more consistent and systematic during fall 2009.

As alluded to earlier, observers were instructed to provide capsule ratings of each lesson they observed. A summary of the capsule ratings showed that MTTI teachers had good pedagogical knowledge: 59% received a high on the rating related to "beginning effective instructional strategy." A more detailed breakdown of the capsule ratings appears in Table 10.

Table 10. Capsule Rating (n = 75, median = 4 or Beginnings of effective instruction - Solid)

	Frequency	Percent
Ineffective Instruction	2	2.7
Elements of effective instruction	2	2.7
Beginning of effective instruction (Low)	5	6.7
Beginnings of effective instruction (Solid)	30	40.0
Beginnings of effective instruction (High)	9	12.0
Accomplished effective instruction	27	36.0

Student Engagement and Cognitive Activity Levels. Student Engagement (SE) and Cognitive activity (CA) levels were also noted during the observations. **SE** was defined as **high** when 80% or more of students were engaged, as low when 80% or more of students were off-task, and as **mixed** otherwise. Of the 903 five-minute intervals, student engagement was noted 779 times. Students were highly engaged 41% of the time, somewhat engaged 48% of the time, and minimally engaged about only 10 % of the time.

Five levels of CA were coded in the observations:
- Level 1 (Receipt of Knowledge) was noted when students were involved in the reception of information;
- Level 2 (Application of Procedural Knowledge) when students applied their knowledge by doing worksheets, practicing problems or building skills;
- Level 3 (Knowledge Representation) when students manipulated information by reorganizing, categorizing or attempting to represent what they learned in a different way;
- Level 4 (Knowledge Construction) when students created new meaning by making connections, generating ideas or solving new problems;
- Level 5 (other) for administrative tasks, interruption etc.

Of the 903 five-minute recordings available, Level 1 and Level 2 were recorded 94% of the time, meaning that students spent most of their time receiving and applying procedural knowledge.

Table 11. Student Engagement and Cognitive Activity Levels

Student Engagement Levels	n observed	Percent	Cognitive Activity Levels	n observed	Percent
Level 1- Low: 80% or more of students were off-task	81	10.4	**Level 1-** Receipt of Knowledge	484	72.8
Level 2- Mixed	375	48.1	**Level 2-** Application of Procedural Knowledge	144	21.7
Level 3- High - 80% or more of students are engaged	323	41.5	**Level 3-** Knowledge Representation	27	4.0
			Level 4- Knowledge Construction	8	1.2
				12	21.8
			Other		
Total	779	100		665	100

Lesson Plans

All teachers submitted two sample lesson plans per observation. These lessons were to be representative of how and what they actually did on a regular basis. The plans were rated by the observers and Lehman College's mathematics education faculty using the Lesson Plan Assessment Rubric (LPAR), which was also designed through the same process as the COP. The rating protocol focused on essentially the same student-centered elements as the observation protocol. Ranging from 1 (not evident) to 4 (clearly evident), the ratings indicated the extent to which the lesson plans "encouraged students to manifest characteristics of students in standards-based classrooms. As with the Observation Protocol, a "Capsule Rating" also provided the planned lesson's effectiveness ranging from 1 (ineffective lesson) to 7 (exemplary lesson).

Results

Critical thinking. Adequate involvement of students' critical, in-depth, higher-order thinking was rated in 49% of the lesson plans (n = 97), whereas substantial involvement of students was noted in only 4% of the lessons.

Table12a. Critical, in-depth, higher order thinking

	Frequency	Percent
student not involved in higher-order thinking	5	5.2
superficial involvement in higher-order thinking	30	30.9
adequate involvement of students	47	48.5
Substantial involvement of students	4	4.1

| Not applicable/don't know | 11 | 100.0 |

Concrete connections to the real world was rated in 20% of the lessons (n = 97). Fifty-nine (60%) indicated no personal or real-world connection. Students' exploration of their own connections to the real world was rated only in 1% of the lessons.

Table 12b. Real-world connections (n = 97)

	Frequency	Percent
no connection to real world	30	30.9
familiar objects presented but no personal connection	29	29.9
public problem concrete connection	19	19.6
Students explore own connections to world	1	1.0
Not applicable/don't know	18	18.6

Work as a learning community. 42% of the lessons (n = 97) indicated that students would share some information; 31% indicated that teachers would direct their classes to debate, discuss, analyze, and/or evaluate the information. Only 5% indicated that students would always work independently.

Table 12c. Work as a learning community (n = 97)

	Frequency	Percent
students always work independently	5	5.2
some sharing of information	41	42.3
teacher directs class to debate, discuss, etc.	30	30.9
Students raise problems they think important	1	1.0
Not applicable/don't know	20	20.6

Meeting different students' needs. 62% of the lessons (n = 95) showed the use of two or three modalities, and 24% the use of at least four modalities. Because use of multiple modalities can accommodate for diversity—disabilities, learning abilities, gender and/or cultural differences—raters had to see the use of more than one presentation modality: kinesthetic/tactile; oral/verbal; written; numerical; and problem resolution through the use of equations, graphs or charts, pictures, tabulations.

Table12d. Different student needs met (n = 95)

	Frequency	Percent
One modality used	8	8.2
Two or three modalities used	60	61.9
lesson presented using four modalities	24	24.7
Not applicable/don't know	3	3.1

Grasp of content. The learning of good, solid content as it connected to broader concepts of the discipline was noted in 74% of the lessons (n = 95).

Table 12e. Learn important concepts for understanding discipline (n = 96)

	Frequency	Percent
basic content, no connection to broader concepts	19	19.6
good, solid content, somewhat connected	72	74.2
good content substantially connected	2	2.1
Not applicable/don't know	3	3.1
Total	96	99.0

In the final rating of each lesson plan, raters had to consider all available information about it, its context and purpose, and use their own judgment of the relative importance of the ratings they made to select a **capsule description** (not an average) that best characterized the quality of the lesson plan and its likely impact on student learning.

Capsule Rating Results. The following represents the results of the capsule ratings of the 96 lesson plans, collected from 28 teachers:
- 4% demonstrated elements of an effective lesson;
- 24% demonstrated the beginnings of an effective lesson (Low);
- 40% demonstrated beginnings of an effective lesson (Solid);
- 9% demonstrated beginnings of an effective lesson (High);
- 22% demonstrated an accomplished, effective lesson.

Table13. Lesson Plans Capsule Rating (n = 96)

	Frequency	Percent
Elements of effective lesson	4	4.1
Beginnings of effective lesson (Low)	23	23.7
Beginnings of Effective lesson (Solid)	39	40.2
Beginnings of effective lesson (High)	9	9.3

Accomplished, effective lesson	21	21.6

Relationship Between Observations and Lessons Plans

The analysis below emanates from 83 observations of 34 teachers, and the collection of 97 lesson plans from 28 of these teachers. Fifteen teachers were observed twice; 17 teachers were observed three times; two teachers were observed once. Two teachers submitted 1 lesson plan, nine teachers 2 lessons, four teachers 3 lessons, four teachers 4 lessons, five teachers 5 lessons, and four teachers 6 lessons. The table below shows the distribution of observations and lesson plans.

Table 14: Frequencies of Lesson Plans submitted and observations

Lesson Plans				Observations		
	Frequency	No of Lessons			Frequency	No of Observations
One lesson	2	2	One Obsv.		2	2
Two lessons	9	18	Two Obsv.		15	30
Three lessons	4	12	Three Obsv.		17	51
four lessons	4	16				
five lessons	5	25				
Six lessons	4	24				
	28	97			34	83
Total	Teachers	Lessons		Total	Teachers	Observations

Results. In spring 2009, we piloted the lesson plan and observation instruments with 20 teachers through four observers and found a correlation at the .05 level between the lesson plans (n = 22) and observations (n = 20) capsule ratings (r = .522, p = .018). In fall 2009, using a different set observers that collected 97 lesson plans and did 75 observations, we also found a correlation at the .01 level (r = ..350, p = .006). Moreover when lesson plans and observation ratings were averaged for each teacher, giving one value per teacher for each measure, the correlation was even more significant (r = .490, p = .008, α < .01). This correlation across observers and across time shows the reliability of the two instruments that were used to determine teachers' ability to teach.

Lesson plans capsule ratings also correlated with teachers' pedagogical knowledge as measured by the Louisville instrument (r = .425, p =.024, at the). Teachers with good pedagogical content knowledge tend to prepare good lessons. There were no correlations between the observation capsule rating and teachers' pedagogical knowledge.

Table 15.

		Spring 2009		Fall 2009	
		Capsule Rating Observation (n = 20)	p	Capsule Rating Observation (n = 60)	p
Spring 2009 (Pilot)	Capsule rating Lesson Plans (n = 22)	**.522***	.018		
Fall 2009	Capsule rating Lesson Plans (n = 96)			.350**	.006

* Significant at the .05 level --   ** significant at the .01 level.

Relationship between Louisville Tests and Lesson Plans

The following correlations were also found: whereas the Capsule Lesson Plan ratings (n =28) correlated with the following **sub-categories** of the **Louisville tests,** the averaged Observation Capsule Rating did not correlate with any of these subsets:

Table 13

	Averaged Capsule rating of Lesson Plans	
Pedagogical Content K/10	r = .425*	p = 0.24
Expressions and Formulas /9	r = .456*	p = .015
Geometry Cognitive Understanding K/10	r = .436*	p = .020
Geometry Pedagogical Content K/10	r =.531 **	p = .004
Geometry Knowledge Total/40	r = .474*	p = .011
Geometry Three-dimensional geometry/11	r =.518 **	p = .005
Geometry Transformational geometry /9	r = .447*	p = .017
Geometry Measurement/9	r = .380*	p = .046
Geometry Sub-categories Total/40	r = .474*	p = .011

* Correlation is significant at the 0.05 level.   ** Correlation is significant at the 0.01 level

Teacher Interviews

Fourteen, randomly-selected teacher-participants were interviewed between July 6 and August 7, 2009. The interviews were based on a semi-structured interview protocol, and took approximately 30 minutes to conduct. Interviewees' responses were digitally recorded and later transcribed. Participants were not identified by name on the transcripts.   The interview schedule contained five sections:

- A-Professional Development;
- B-Teacher Leadership;
- C-Professional Goals;
- D-Math Content Knowledge;
- E-Perceptions of the MTTI Program.

Results, A-Professional Development.

Prior to MTTI. Teachers were asked what professional development they had received in the year prior to starting the MTTI program. One respondent said that they had not taught math in the previous year, and consequently did not receive any professional development in math in that year. Of those participants who taught math in the year prior to the MTTI program, six said they received extensive professional development. Of these, four said they were providing, or helping to provide, the math professional development in their school. Most of the six respondents said that they were part of math professional development networks, either with other local schools or schools district wide and, in one case, nation-wide. Two interviewees said that they were primarily responsible for providing technical support and professional development rather than math professional development. Five said that they had participated in little or no professional development in math in the year prior to starting the MTTI program.

When asked to estimate the influence prior professional development had on their math teaching, using a hypothetical 11-point scale (0=no influence; 10=total influence), the average rating for 13 participants was 6.08 (minimum = 2; maximum = 8).

MTTI participation. When asked, using the same scale, to estimate the influence MTTI participation had had on their math teaching, the average rating for 13 participants was 3.17 (minimum = 0; maximum = 8). The primary reason for the relatively low ratings was that they had only been participating in the MTTI program for a semester; there had not been enough time for professional development to influence their practice.

Results, B-Teacher Leadership.

Prior to MTTI. Five respondents said they had had little or no experience in a teacher-leadership role in math prior to participating in the MTTI program. The other nine teachers mentioned many and various examples

of teacher-leadership roles they played in their schools. These included: running seminars for fellow teachers in their schools; contributing to staff development programs outside the school; networking with other schools; and attending national conferences. In addition, three teachers mentioned being involved in the technical development of fellow teachers.

MTTI participation. There was no indication that MTTI participation had had any effect on participants' teacher-leadership roles. Again, because of the relatively short period of time in the MTTI program, participants have yet to engage in components of the program specifically designed to enhance teacher-leadership skills.

Participants were asked to indicate the degree of support they had from their school's administration in relation to developing their role as a teacher leader. They were asked to use a hypothetical scale (0=no support;10=total support). The average rating for 10 participants was 6.5 (minimum = 0; maximum = 10).

Results, C-Professional Goals.

Pedagogical goals. Teachers' pedagogical goals varied considerably. One said that their goal was to have their students pass high-stakes test. Another said that they wanted to meet the needs of their high-achieving students. A third said they wanted to improve their own math content knowledge. While a different respondent said they wanted eventually to be able to teach AP calculus and AP statistics. One said that they wanted to be better at classroom management, and in a similar vein, one said they wanted to improve their time management. Two wanted to work on increasing differentiated learning in their classes. Two said they wanted to eventually become assistant principals or principals.

Participants were asked to indicate how confident they were that they could teach math to all their students. They were asked to use a hypothetical scale (0=not at all confident; 10=totally confident). The average rating for 14 participants was 8.1 (minimum = 5; maximum = 10).

Teacher-leadership goals. Most participants indicated that they wanted to keep themselves informed about math pedagogy and improve their content knowledge. This was in order to inform other teachers at their school and help them to develop a deeper understanding of math. They also wanted to inform other teachers about critical parts of the

curriculum. Three teachers said their goal was to incorporate technology and have other teachers be more comfortable with technology. Respondents were asked to indicate how confident they were that they could play a teacher-leadership role in their school. They were asked to use a hypothetical scale (0=not at all confident; 10=totally confident). The average rating for 14 participants was 7.6 (minimum = 2; maximum = 10).

Results, D-Math Content Knowledge. All respondents agreed to some extent with the view that increasing their own math content knowledge could lead to an improvement in their students' understanding of math. However, some said that they had difficulty in dealing with some of the more advanced concepts in both fundamental math and geometry that they had been exposed to in the first semester of the MTTI program. These teachers tended to say that they were confident in their knowledge of the math that they had to teach to their students, but not so much in more advanced areas.

Respondents were asked to indicate how confident they were generally in their own math content knowledge. They were asked to use a hypothetical scale (0=not at all confident; 10=totally confident). The average rating for 12 participants was 7.3 (minimum = 5; maximum = 10).

Surprisingly, there was no significant correlation between teachers' ratings of their confidence in their own math content knowledge and their confidence that they could teach math to all their students ($r = -.026$, $p = .937$). In addition, there was no significant correlation between teachers' ratings of their confidence in their own math content knowledge and their performance on the combined Louisville tests of Algebra and Geometry ($r = .304$, $p = .336$). There was also no significant correlation between teachers' ratings of their confidence in their own math content knowledge and scores on those items that assess pedagogical knowledge on the combined Louisville tests of Algebra and Geometry ($r = .327$, $p = .299$).

A similar pattern of non-significant correlations emerged when teachers' confidence in being able to teach math to all their students was related to their performance on the combined Louisville tests of Algebra and Geometry ($r = .309$, $p = .282$), and scores on those items that assess pedagogical knowledge on the combined Louisville tests of Algebra and Geometry ($r = .257$ $p = .375$).

Results, E-Perceptions of the MTTI Program.

Elements Found to Be Beneficial. Four respondents said they found the content of the Fundamental Math or Geometry courses beneficial. This was generally because the courses stretched them and deepened their understanding of math. One of these respondents said that they had tried to implement some of what they learned in the MTTI course in their classroom.

On the other hand, four teachers indicated that they found little of benefit, and that they had found little of what they had been taught in the MTTI course relevant to their own teaching.   They also found the content too difficult for them to cope with.

Another four teachers said that they found the collaboration and discussion with fellow teachers very useful. Participants tended to help each other, and discuss pedagogy with one another. One person responded that the tutoring support was very good.

Ways in Which the MTTI Program Could Be Improved. Consistent with the findings reported in the previous section, most respondents said that the MTTI program could be improved by making the content in the courses more relevant and applicable to their classroom teaching. In addition, they thought that it would be useful to introduce a course related to teacher-leadership early in the program. This was seen as particularly important because developing teacher-leaders is one of the primary goals of the program.