ETS®

Math Out of the Box Curriculum in Lawrence Township Public Schools

First Year Report

August 23, 2006

Phil Handwerk
Rosalea Courtney
Lorin Romeo

Center for Foundational & Validity Research
Educational Testing Service

Listening.
Learning.
Leading.

EXECUTIVE SUMMARY

This report for the Lawrence Township Public Schools (LTPS) focuses on the first year of Math Out of the Box (MTB) curriculum as implemented in the township public schools.

- Most of the stakeholders interviewed regarded the pilot of MTB as an opportunity to document how teachers changed in their knowledge of mathematics and use of inquiry-based teaching strategies.
    - The most salient dimension mentioned by stakeholders was the enthusiasm of the teachers for MTB. This has encouraged more teachers to become involved in the pilot and seems to have had a positive impact on students and parents.
    - One concern expressed primarily by LTPS administrators was uncertainty about what the MTB program would look like in its final form.
- The professional development sessions met the standards for high quality inquiry-based pedagogical training. Teacher reaction to the training was very positive, although some mentioned a need for additional kit-specific training.
- Classroom observations found teachers to be successful in implementing the MTB curriculum.
    - In most lessons, the mathematics was standards-based, appropriate, and challenging. Typically, students had opportunities to communicate their understanding through discussion and/or writing. The lessons allowed students to apply their understanding in activities that went beyond drill and practice.
    - Students were actively engaged in math learning throughout the lesson and were often given opportunity to work together in small groups in a collaborative way.
    - Teachers used a variety of questions – both higher order and factual recall, some of which led to open discussion which pulled the students into the analysis or brainstorming in a way that one-question-one-answer sequences cannot.
    - Since most small group work required students to solve problems together or construct lists of attributes, the conversations appeared to be moving them towards greater understanding or comfort with the mathematical concepts.
- Overall, the mini-assessments developed in the first year of the evaluation were successful in achieving their primary goal of creating items to be used in the pre-/post-assessments to be given to all 3rd through 5th grade students in 2006-07.
    - A total of 132 multiple-choice items and 36 open-ended items, were piloted with 245 students in the 2005-06 academic year, covering the threes MTB strands (Algebra: Patterns, Algebra: Data Analysis, and Geometry).
    - These items were found to be measuring a similar construct as that measured by the 2006 NJ ASK mathematics subject test, thereby justifying their future use.
    - Teachers who participated in the scoring of the mini-assessments contributed to the design of the 2006-07 assessments through feedback on specific items. Additionally, some identified a need for further work with their students on open-ended questions.

Math Out of the Box Curriculum in Lawrence Township Public Schools

First Year Report

This report for the Lawrence Township Public Schools (LTPS) focuses on the first year of Math Out of the Box (MTB) curriculum as implemented in the township public schools. It is to provide LTPS with feedback on the implementation of the program for their internal use. First, we will give a brief review of the curriculum and the purposes and history of the evaluation to date. Next we will discuss findings from the qualitative research followed by the quantitative analysis of assessment data. Finally, we outline future steps and goals for the evaluation.

1. INTRODUCTION

In 2005 Educational Testing Service (ETS) was invited by the DuPont Center for Collaborative Research in Education (DuPont) to participate in the implementation of MTB in LTPS. ETS agreed to conduct an evaluation and provide accommodations for the professional development workshops. This evaluation was conducted in kindergarten through 5^{th} grades in the Ben Franklin Elementary, Eldridge Park Elementary, Lawrenceville Elementary, Slackwood Elementary, and the Lawrence Intermediate School.

The first group of teachers (Cohort 1) participated in professional development in the spring of 2005. There were 18 teachers in the original group that volunteered for MTB, although reassignments reduced this group to 15 in the 2005-06 school year. The Cohort 1 teachers taught both the Developing Algebraic Thinking (Algebra) and Developing Geometric Logic (Geometry) MTB strands during the year. In February 2006, the second group of volunteer teachers (Cohort 2) was introduced to MTB and began teaching the Algebra strand in the spring. There are currently 18 teachers in Cohort 2.

The evaluation design was derived from strategic planning among the project's partners (ETS, LTPS, DuPont, and Clemson) and is moving forward on schedule. Table 1 presents a summary of the events beginning in the Spring of 2005 through the Spring of 2006.

Table 1. Calendar of MTB Evaluation 2005-06

Date	Event
Spring 2005	• Cohort 1 teachers participated in professional development/training workshops.
Fall 2005	• Cohort 1 teachers taught Algebra: Patterns strand. • Students completed Algebra Patterns pre- and post-assessments.
Winter 2005-06	• Cohort 1 and 2 teachers participated in professional development/training workshops. • Cohort 1 teachers scored the Algebra Patterns assessments.
Spring 2006	• Cohort 1 and 2 teachers participated in professional development/training workshops. • Cohort 1 teachers taught the Geometry strand. • Cohort 2 teachers taught Algebra: Data Analysis strand. • Students of Cohort 1 teachers completed the Geometry pre- and post-assessments. • Students of Cohort 2 teachers completed the Algebra: Data Analysis pre- and post-assessments. • Cohort 1 teachers scored the Geometry assessments. • Cohort 2 teachers scored the Algebra: Data Analysis assessments.

a. Purpose of qualitative evaluation

The primary purpose of the qualitative evaluation of this project was to provide periodic feedback to the district about teachers' responses to the materials and training. Two memos were provided to the district in 2005 and 2006 reviewing teachers' feedback during professional development reflection and individual interviews.

In this report, one major focus will be a description of classroom observations during MTB lessons. These algebra and geometry lessons were taught by Cohort 1 teachers in grades K-5 and observed during the winter and spring of the 2005-06 school year.

Another qualitative evaluation goal was to listen to representatives from all the groups of stakeholders about their vision for the pilot project and its evaluation. ETS staff interviewed 12 stakeholders, including all of the participating principals, during the spring of 2006. The purpose of these interviews was to document the understandings about the goals and outcomes from all of those vested in this pilot.

b. Purpose of quantitative evaluation

The quantitative portion of the 2005-06 evaluation was intended to serve as a means of piloting assessment items that will be used in the pre-/post-assessments to be given to all 3rd, 4th, and 5th grade students at the beginning and end of the 2006-07 academic year.

In this past year, students in MTB classrooms completed 15 minute mini-assessments before and after completing the Algebra: Patterns, Algebra: Data Analysis, and Geometry MTB strands. Each mini-assessment was comprised of multiple-choice items and open-ended items. MTB teachers then participated in a professional development day, lead by ETS Assessment Developers in which all completed mini-assessments were discussed, graded, and suggestions made to improve items. Graded assessments were then analyzed, item by item, to review the performance of the items, get an understanding of their relative difficulty, and obtain a general idea as to overall student performance.

Because the assessments themselves only contained 8 to 12 items, and there were relatively small numbers of students taking each assessment (ranging from 13 students to 65 students), little emphasis should be placed on the outcomes of these piloted assessments. Furthermore, when constructing the pre- and post-assessments for each grade and strand, the goal was to pilot as many different types of questions covering as much material as possible. Therefore, the pre- and post-assessments were not designed to be equivalent forms, which then does not allow for a fair comparison between students' performance before versus after receiving the MTB instruction.

2. MTB PROGRAM COMPONENTS

This section will describe briefly the MTB curriculum and its history. Additionally, there will be a recap of the MTB training given to the LTPS teachers.

a. Description curriculum for MTB

Two MTB modules were implemented in LTPS during the 2005-06 academic year. The Algebra strand was available to teachers in the spring of 2005 and the Geometry strand was launched in the spring of 2006. The modules, marketed by Carolina Biological Supply Company, consist of a large box of materials sufficient for 20-25 students and a MTB Teacher Guide. Within the box are a variety of materials such as sets of colored blocks representing different 3-dimensional shapes, 2-dimensional tiled shapes, white boards, calculators, and protractors.

The MTB Teacher Guide includes quizzes and tests for the module and reproducible practice pages. LTPS agreed with MTB to participate in the piloting of these modules and Teacher Guides while they are still in draft form. Feedback from teachers participating in the pilot will contribute to further refinement of the Teacher Guide.

Mathematical topics in MTB are chosen and sequenced vertically between the grades and horizontally within each grade level, resulting in a connected curriculum. The curriculum is designed so that students will develop the ability to make the following mathematical connections:

- Among mathematical ideas, facts, and procedures,
- Between mathematical ideas and other disciplines, and
- To their own environment.

In the MTB curriculum, students engage in an inquiry-based learning cycle with special emphasis on developing reasoning, problem-solving, and higher order thinking skills. Further, helping students to develop the ability to communicate about mathematics is another key feature of this curriculum. Communication can take the form of discussion, questioning, reflection, and writing to ensure that meaningful mathematical thinking occurs.

b. Description of the training

Prior to implementing the MTB curriculum in their classrooms, teachers participated in a full day workshop introducing them to the materials in their kits. They then participated in a second training midway through the kit and a follow-up reflection session after completing the kit.

Table 2. Timing of MTB Training by Teacher Cohort

Teacher Cohort	Algebra Strand	Geometry Strand
Cohort 1	February 2005, April 2005, and May 2005	March 2006, April 2006, and May 2006
Cohort 2	March 2006, April 2006, and May 2006	

Teachers across the grades K-5 attended the training for their Cohort. The cross-grade participation necessitated training more general the than typical "kit-based" training that is usually grade specific. At the first and second training workshops, sample activities from the kindergarten, 3rd grade, and 5th grade kits were introduced to all teachers.

The workshop facilitators presented training that was inquiry-based and introduced teachers to the major components of the program including use of materials and tools for communication. The professional development sessions were met the standards for high quality inquiry-based pedagogical training. Teacher reaction to the training was very positive, although some mentioned a need for additional kit-specific training.

3. QUALITATIVE EVALUATION

This section will review the findings from qualitative evaluations conducted during the 2005-06 academic year. First, there will be a discussion of the stakeholder interviews, followed by a review of the classroom observations.

a. Stakeholder interviews

ETS research staff interviewed MTB developers from Clemson University, DuPont foundation representatives, ETS corporate leadership, and the Lawrence Township Public School administration including all five principals. Although some questions overlapped between the groups, other questions were more audience-specific. For example, principals who had opportunity to observe MTB classrooms were asked to share those impressions while Clemson staff were asked questions regarding developing the curriculum.

In this section of the report, we will be looking first at questions that can be looked at across groups of stakeholders. These questions include:

- How did the partnership for the LTPS/MTB pilot develop?
- What are your expectations for this 3-year pilot?
- How would you evaluate the first year of the pilot partnership so far?
- Do you have any concerns for the future?

The history of LTPS' participations in MTB was most fully described by LTPS district administration and DuPont representatives. The LTPS staff was looking for an alternative curriculum in math that would be more effective than the adopted textbook series. This search led to DuPont and then to Clemson and Carolina Biological.

Most of the stakeholders interviewed regarded the LTPS pilot as an opportunity to document how teachers changed in their knowledge of mathematics and use of inquiry-based teaching strategies. They were also interested in the impact that MTB would have on students as evidenced in math scores and attitudes.

As the stakeholders evaluated the LTPS project to date, the most salient dimension was the enthusiasm of the teachers for the curriculum. As one principal said, "I didn't expect to see that much passion." Teacher enthusiasm was also evident during the professional development sessions. Several stakeholders noted that more teachers are asking to become involved in the pilot and this excitement seems to have had an impact on students and parents as well.

One concern that was expressed primarily by district and building administrators involved uncertainty about what the MTB program would look like in its final form. Since the modules have been launched one-by-one with only two out of five currently implemented in the district, there was concern whether the entire program will stand alone or serve as supplemental material to some other program. There were also concerns about the changes in district administration and uncertainty about what impact that would have for the MTB pilot.

b. Classroom observations

An important component of qualitative evaluation of a new curriculum or instructional program is determining the extent to which teachers successfully implement the new curriculum. In the spring 2006 Interim Memo, ETS summarized teachers' ratings of the five MTB classroom dimensions developed by Clemson University. In the Dimension Survey, teachers self-reported how likely they were to use inquiry and traditional strategies when teaching MTB. The teacher ratings were not scaled, but were rated with a plus or minus for each subheading.

To provide another view on teacher implementation of the MTB curriculum, Cohort 1 teachers were observed teaching Algebra and Geometry modules. The classroom observations were voluntary. All 16 of the teachers were contacted of whom 8 scheduled observations. A total of 14 MTB Algebra and Geometry lessons in five grades were observed. See Table 3.

Table 3. Number of Lessons Observed by Grade and MTB Strand

	Algebra	Geometry
Kindergarten	2	2
1st Grade	1	1
2nd Grade	1	0
4th Grade	2	2
5th Grade	1	2

Narrative notes were made while the lesson was observed and these in turn were coded along four areas. In coding the lessons, a framework was used that modified a coding scheme developed by Horizon Research,[1] but that also incorporated elements from the MTB dimensions. The coding scheme focused on four areas:

- Design and Implementation of the Lesson
- Math as a Discipline
- Classroom Interaction and Environment
- Questioning and Classroom Talk

Individual lessons were coded separately for each of these four areas on a five point scale (5 is high and 1 is low) that differentiated the extent to which each of the components was observed during the lesson and whether or not that area was likely to enhance student learning. The four areas and their component statements represent agreed upon best practice in inquiry instruction and are in part the focus of MTB instruction. For many reasons, not all components are possible in every lesson, so the absence of some components would not adversely affect the overall rating unless its inclusion would seem to strengthened student understanding. A discussion of each of the areas is reported below along with overall scores reported for all 14 lessons.

[1]Horizon Research, Inc. (2000). *Local Systemic Change: 2000-2001 Core Evaluation Data Collection Manual.* Chapel Hill, NC: Author.

Design and Implementation of the Lesson

The Design and Implementation of the Lesson focuses on the overall cohesion of the lesson. This goes beyond the lesson plan to include how that design was implemented. Because the kit includes materials for individual students or groups of students, the review of materials seems most appropriately placed in this section. Specifically we looked at whether or not the materials were used to support student investigation leading to greater understanding of math concepts and not for demonstration or illustration purposes.

Other subheadings under the Design and Implementation of the Lesson include the variety of avenues provided to students to understand the content, scaffolding of concepts, and framing the lesson at the beginning with discussion of prior learning and at the end with some type of wrap-up. Teachers could provide a variety of avenues (large group discussion, small group investigations, and individual responses) which may or may not be designed to scaffold a concept from simpler ideas to more complex understanding. Framing of the lesson is important because it provides cognitive organization of the content to students in a way that may not otherwise be clear. However, many of the MTB lessons spanned two or three days (or several math periods) so a wrap-up was often not part of the observed lesson.

The average score for this section across the 14 observations completed in Cohort 1 classrooms was 3.9 with most of the lessons falling in the 4 or 5 level for this section. Teachers seem to be adhering fairly closely to the "script" provided by the MTB Teacher Guide. In a few lessons the pathways or strategies for understanding were limited. For example, in two lessons there was only whole class teacher-directed activity throughout the lesson with no opportunity for small group work or for individual writing.

In most lessons, teachers introduced the lesson and led a whole class discussion that set the stage for small group and individual activities. To illustrate, in a first grade geometry lesson, the class discussed the properties of cubes and cylinders covered in an earlier lesson. There was also a discussion about the vocabulary (e.g., "analyze," "attribute," "faces," "edges") and the students were asked to identify the different geometry shapes included in the kit. A square pyramid was placed in the center of the circle and each child was given a block. If they were able to match their block to the pyramid on some attribute, they placed it in the center. They moved to small groups and each group was given a set of shapes (prisms, cones, spheres, etc.) and working together they developed a list of attributes that described all of their blocks. This activity concluded with sharing across groups and teacher questioning that led to comparisons between different shape categories. After this, an individual activity involving sorting and categorizing by attributes provided information about student learning.

Math as a Discipline

MTB curriculum is designed to be standards-based and to focus learning on conceptual understanding rather than drill and practice. There were four subheadings under Math as a Discipline that were coded in this section. The first was whether the content was standards-based, the second addressed opportunities provided in the lesson for students to communicate their understanding through discussion and/or writing. The third looked specifically at the activities in the lesson (ie., Were students engaged in activities that encouraged higher order thinking?). The last area focused on whether or not connections were made to other disciplines or to real-world applications.

The average score for the fourteen lessons was 4.1 for this section. In most lessons the mathematics was standards-based, appropriate, and challenging. Typically, students had opportunities to communicate their understanding through discussion and/or writing and in some instances through both modalities. The lessons allowed students to apply their understanding in activities and problems that went beyond drill and practice. This was true across all the grades including kindergarten. Although opportunities to make real world connections seemed available, in few lessons was this connection explicit.

Classroom Interaction and Environment

In more traditional classroom settings, students have little opportunity to interact with each other around math concepts. Frequently, the lesson is teacher-directed in a whole class setting. Students are seldom seen interacting with each other around materials and problems, since the focus of the lesson is the teacher.

Four subheadings were coded in the Classroom Interaction and Environment section. The first focused on the extent to which a collaborative climate was evident during the lesson. This was demonstrated by exchanges initiated by students as well as the teacher that focused on math and were respectful of ideas of all. The second subheading looked at student engagement in the mathematical ideas in the lesson. At the highest level of rating for this subheading, all of the students would be engaged in math learning all of the time. The third subheading focused on the extent to which solving problems or completing group activities was worked out by student collaboration. And last, did students provide feedback to one another in a constructive manner or evaluate their own work.

The overall average rating for this section was 4.1. In most observations, students were actively engaged in math learning throughout the lesson. Students were often given opportunity to work together in small groups in a collaborative way. In a 4th grade classroom at different points throughout the lesson the teacher asked a question and told the students to talk it over with a neighbor. Then students reported out to the class what their neighbor had said. In a 5th grade geometry lesson, several students were engaged in debates about what constituted symmetry. The students used the materials from the kit as well as math textbooks to make their points. However, there were a few lessons where the teacher directed the lesson either in a whole class setting or the supervision of activities completed individually and in those lessons students had less opportunity to interact with one another around the content of the lesson.

Questioning and Classroom Talk

Inquiry has a lot to do with the kinds of questions that are posed to (or from) students and how the Q & A exchanges are played out. Questions that allow more than one answer and classroom exchanges that probe for deeper understanding as ideas bounce off of each other are more likely to lead to higher order thinking. This is in contrast to teacher questions that imply "guess what I'm thinking."

Although this area overlaps somewhat with Classroom Interaction and Environment, there is a difference in focus. The Classroom Interaction and Environment code focuses more on the climate of the classroom and the opportunities for collaboration. In Questioning and Classroom Talk, the focus is more on the nature of the exchanges especially the questioning sequences between teacher-student and student-student.

The first subheading looks at what types of questions teachers ask whether or not higher level questioning was included. It was understood that lower level, factual questions play a role in classroom interaction. The issue here was to what extent higher level questions were also being used. In the second subheading, the focus was more interactive (i.e., Is there discussion that allows students to bounce ideas off of each other? Does the teacher go beyond a simple Q&A cycle to probe for deeper understanding?). Also included in this section was the focus of student-to-student talk as it unfolded during small group work. To what extent did the conversations among students foster greater understanding of the mathematics concepts? The last subheading looked at the use of teacher feedback that built understanding and went beyond answers simply being categorized as right or wrong.

The Questioning and Classroom Talk area had an overall rating of 3.5. Many of the teachers used a variety of questions – both higher order and factual recall. Many of the questions used in the lessons observed came from the MTB Teacher Guide. The questions in these manuals represent a mixture of higher order questioning and factual questions especially around vocabulary. In the observations, questions like "What is another word for square? It begins with c…" were frequently observed.

Some teacher questions led to open discussion that pulled the students into the analysis or brainstorming in a way that one-question-one-answer sequences cannot. The following is an example of the teacher/student interaction in which the teacher showed students different shapes on at a time:

> Teacher: What do you notice about this pyramid?
> Student 1: Has a square base.
> Student 2: Line on the bottom divides it.
> Student 3: It has 4 triangles.
> Student 4: Has 5 faces.
> Student 5: Wouldn't be good for a building.
> Student 6: Could be like a roof.

The opportunities for students to exchange ideas in small groups varied. In some lessons there was no small group work and all of the discussion was teacher-led. In classes where small group work was part of the lesson, the question for this subtopic was to what extent the small group discussion was moving students toward greater understanding. Since most small group work required students to solve problems together or construct lists of attributes, the conversations appeared to be moving them towards greater understanding or comfort with the mathematical concepts.

Algebra Lessons vs. Geometry Lessons

Six of the Cohort 1 teachers were observed teaching both Algebra (fall 2005) and Geometry (spring 2006) MTB lessons. Although the Geometry strand was a new module for them (the Teacher Guide were still in draft form), average scores for Geometry were higher for Lesson Design and Implementation, Classroom Interaction/Environment, and Questioning and Classroom Talk areas than was true for the Algebra lessons. Table 4 presents a summary of the averages of the Algebra and Geometry lessons for the six teachers who were observed teaching both strands. The overall averages summarized for the 12 lessons are also included.

Table 4. Average Subscores in Algebra and Geometry MTB Strands for 6 Teachers Observed During Both Strands

	Algebra (n=6)	Geometry (n=6)	Overall (n=12)
Lesson Design	3.8	4.1	4.0
Math as a Discipline	4.1	4.1	4.1
Classroom Interaction/Environment	3.8	4.5	4.1
Questioning and Classroom Talk	3.0	3.7	3.3

Clearly the small numbers of observations shown in the table above makes it difficult to draw conclusions. Several explanations of the differences between the Algebra observations and the Geometry observations are possible. It may be that the teachers are becoming more comfortable with the curriculum or MTB Geometry is easier to implement. It is also possible that the Geometry module lends itself to inquiry approaches more easily than the Algebra module. We will continue to observe Cohort 1 math lessons in the 2006-07 school year to see if familiarity with the materials and activities, as well as future professional development, would cause the trend to continue.

In general, the teachers in Cohort 1 were successful in implementing the MTB curriculum in their classrooms. They are already beginning to plan extensions in terms of introducing math trade books for children into the lessons. In some grades, teachers have begun thinking about how to extend and/or collapse lessons as needed. Input from teachers experienced in inquiry-based instruction should provide the curriculum developers with the level of feedback that will add to the further refinement of the curriculum.

4. QUANTITATIVE EVALUATION

This section will describe the development of the 18 mini-assessments used during the 2005-06 academic year in the MTB classrooms. It will then report on how these mini-assessments were scored and the information gained from the scoring.

a. Development of the mini pre-/post-assessments

In order for this evaluation's student assessments to be aligned with the learning experience, the assessment exercises must focus on the student's ability to make connections, reason mathematically and solve mathematics problems, and communicate his or her thinking. These assessment characteristics are highly desirable learning outcomes for any student who studies mathematics, regardless of the curriculum to which the student is exposed. Therefore, the assessments were developed in such a way so as to maximize the information gained relative to these characteristics for both MTB and non-MTB students in LTPS.

Two mini-assessments were developed for each of the strands in each grade, for a total of 18 mini-assessments. In almost every grade, for every strand, a pre-assessment was given prior to the teacher using the MTB curriculum. After the last lesson in the strand was taught, a post-assessment was administered. The one exception was in the 5^{th} grade Geometry strand where no pre-assessments were given, but two versions of the post-assessment were administered. See Table 5.

Table 5. Summary of Mini-Assessment Administration

Grade	MTB Strand	Type of Assessment (Admin. Date)	Cohort
3^{rd}	Algebra: Patterns	Pre (Fall '05); Post (Spring '06)	1
	Algebra: Data Analysis	Pre (Spring '06); Post (Spring '06)	2
	Geometry	Pre (Spring '06); Post (Spring '06)	1
4^{th}	Algebra: Patterns	Pre (Fall '05); Post (Spring '06)	1
	Algebra: Data Analysis	Pre (Spring '06); Post (Spring '06)	2
	Geometry	Pre (Spring '06); Post (Spring '06)	1
5^{th}	Algebra: Patterns	Pre (Fall '05); Post (Spring '06)	1
	Algebra: Data Analysis	Pre (Spring '06); Post (Spring '06)	2
	Geometry	Post (2 mini-assessments, Spring '06)	1

Each of the 15 minute mini-assessments included a mix of questions. This mix was comprised of 6-10, multiple-choice questions and two open-ended questions. Some open-ended questions required brief interpretations or the student to demonstrate more in-depth analysis, reasoning, and communication. The multiple-choice questions were clustered together in each mini-assessment. The extended, open-ended questions appeared last in each mini-assessment. This was done to ensure that students would have enough to time to answer as many questions as possible. The number of questions in each mini-assessment was determined in order to strike a balance between the need to obtain as much information on student understanding as possible and the practical consideration of limited available classroom time. Each mini-assessment was 15 minutes in length.

The content coverage of the assessments reflects the topics and the extent to which each topic is emphasized in both the MTB and the LTPS curricula. Since these assessments can only provide a very focused snapshot of students' reasoning, problem-solving, and communication capabilities within the 15-minute time frame, the assessment is not able to provide information on the development of students' basic skills in mathematics.

b. Scoring of assessments

In addition to MTB training, three professional development days were held at ETS with LTPS teachers of MTB to discuss assessment in general, open-ended item development in particular, and to grade the mini-assessments they administered throughout the year. The professional development was lead by the ETS Assessment Developers who designed the assessments used in this evaluation.

Teachers were organized by their cohort then by grade to discuss and score the assessments. Using the rubrics developed for each constructed response item, the Assessment Developers worked with each group of teachers individually to discuss, interpret, and evaluate how the items perform. Teachers were not given their own students' mini-assessments rather they scored the mini-assessments of other teachers within their cohort and grade.

As a result of participation in scoring, some teachers felt need to further reinforce that their students be more thorough in replying to open-ended questions. The issue of writing "too much" in a response to an open-ended standardized question was also discussed as was how to handle that information depending upon whether it was extraneous or incorrect. Teachers observed that there was less MTB terminology than expected on the assessments. (This was by design, to ensure that the items would be accessible to the non-MTB students for next year's study.) Some teachers commented that, based on the student responses to the open-ended questions, that the students were motivated and confident. There were some comments that focused on how closely the content of the items paralleled the MTB content. Specifically, a couple of questions that required the students to apply MTB knowledge directly and then apply it in a somewhat different way prompted one teacher to suggest that a more scaffolded approach to help students transition from one type of application to another might be helpful.

Teachers during the scoring process also contributed to the redesign of the assessments to be given in the following year. Suggestions were made to the Assessment Developers to regarding the use of graphics. For example, they recommended making more prominent keys for pictographs on specific items and providing horizontal grid lines on graphs to make it easier to use. Also it was recommended to place in some item stems the more explicit directions on using labels and titles and encouraging students to "explain how you found your answer." Teachers encouraged the re-ordering of item keys (the correct choice) and distractors (the incorrect options) on specific items to make it less likely a student would find the correct key for the wrong reasons.

c. Results

The piloting of the 18 mini-assessments during the 2005-06 school year was a considerable undertaking. A total of 245 students completed one or more of the assessments covering the three MTB strands. These mini-assessments contained 132 unique multiple-choice items and 36 unique open-ended items. Thirteen teachers from the three grades in the five LTPS schools administered the assessments in conjunction with using the MTB curriculum.

Correlations were calculated looking at the strength of the linear relationship between students' scores on the post-assessments and their 2006 NJ ASK mathematics subject test scores. If the items on the post-assessments are measuring similar constructs as those given in the NJ ASK mathematics subject test, one would expect to see a positive relationship between the two instruments, thus providing some evidence of the validity of using the assessments developed for this evaluation. Post-assessments were chosen for this analysis since they were the assessments given to the students closest to the time of the administration of NJ ASK and after the students have been exposed to the MTB curriculum. Only students in grades 3 and 4 take NJ ASK, therefore only those 3rd and 4th grade students who took both the NJ ASK and the post-assessments were included in the analysis.

Most of the post-assessments showed evidence of being moderately to strongly correlated with the NJ ASK mathematics subject test scores. Two post-assessments failed to provide statistically significant evidence of a non-zero correlation. However, these correlations may have been depressed by small sample sizes, the influence of a few outliers, and the fact that the NJ ASK score to which the specific post-assessments were being correlated is an overall math score. Although NJ ASK mathematics subject test is comprised of several content areas, including "geometry and measurement," "patterns and algebra," and "data analysis, probability, and discrete math," only the overall mathematics score was available. Still, there is some evidence that the post-assessments were measuring a similar construct as that measured by the 2006 NJ ASK mathematics subject test. See Table 6.

Table 6. Correlation of MTB Strands Post-Assessment with 3rd & 4th Graders' 2006 NJASK Math Scaled Score

		Correlation (n=)	
Grade	Algebra: Patterns	Algebra: Data Analysis	Geometry
3rd	.25 (29)	.66* (52)	.54* (32)
4th	.50* (40)	.29 (23)	.56* (42)

*p <.05

Overall, the mini-assessments were successful in achieving their primary goal. The teachers and students provided valuable feedback to the Assessment Developers. The results from this quantitative evaluation will allow for the large, pre/post-assessments for the 2006-07 academic year to be much stronger.

d. Limitations

The purpose of the pre- and post-assessments used in 2005-06 was to serve as a means of piloting assessment items that will be used in the 2006-07 academic year. The 2005-06 pre- and post-assessments were not designed to be equivalent forms. Therefore, a comparison could not be made between how students performed on the assessments prior to receiving the MTB instruction and how they performed on tests after the instruction. In addition, the small sample sizes and the fact that some pre-assessments were administered after the teachers had already begun a few MTB lessons, prevents interpretation of scores on the assessments. Because of these factors, the impact of MTB on these students' content knowledge could not be assessed. In addition, the small sample sizes and numbers of items per assessment limited some of the typical psychometric evaluations done on assessments (e.g., reliability). These analyses will be conducted on the larger assessments to be given to all 3rd-5th grade students during the 2006-07 academic year.

5. FUTURE PLANS

a. Plans for year 2006-07 classroom observation/student focus groups

We will continue to observe MTB lessons in Cohort 1 classrooms during the 2006-07 school year. Teachers are already experienced in using the Algebra and Geometry strands and will be adopting the new Measurement strand in the spring.

During the classroom observations, students were actively engaged in the math learning. The impact of MTB on students' attitudes toward mathematics will be investigated more directly through focus groups.

b. Plans for the assessment of all 3-5th grade students during 2006-07

After reviewing individual item performance, content coverage, and the feedback from teachers, ETS will develop new assessments to be administered in the 2006-07 academic year to all students in grades 3, 4, and 5. There will be one assessment per grade to be given to all students in that grade at the beginning of the academic year. Then again in the spring, one assessment per grade will be given to all students in that grade. Non-LTPS elementary mathematics teachers will be lead by ETS in the scoring of all assessments after their administration in the spring of 2007.

Both the pre- and post-year assessments will be designed to be equivalent forms and will cover all three of the MTB strands. These assessments will have two sections per form – a 25-minute section consisting of 18 multiple-choice questions and a 20-minute section consisting of 3 open-ended questions.

By administering to all 3rd through 5th grade students, a comparison can be made of the performance on the MTB content by students who were exposed to the MTB curriculum throughout the year, and those who were not. Additionally, if 2007 NJ ASK scores are available, comparisons may also be made of how the MTB students scored on that test in comparison to the non-MTB students.