Journal of Urban Mathematics Education
July 2010, Vol. 3, No. 1, pp. 27–56

The Mathematics Learning Discourse Project: Fostering Higher Order Thinking and Academic Language in Urban Mathematics Classrooms

Megan E. Staples
University of Connecticut

Mary P. Truxaw
University of Connecticut

In this article, the authors report results from a small-scale study of the Mathematics Learning Discourse (MLD) project that aimed to affect change in urban mathematics classrooms. The project focused on enhancing students' understanding of mathematics through an emphasis on classroom discourse and higher order thinking. Four teachers participated in a 3-day summer course and yearlong collaboration that was organized around three principles for supporting a learning discourse in their respective classrooms: appropriate and effective development of students' academic language, student engagement in mathematical practices of justification and collective argumentation, and access for all students to rigorous mathematics. The authors discuss the research base for the MLD program, its implementation, and its effect—as well as promise—by analyzing student scores on pre- and post-assessments both for mathematical performance and for the development of students' proficiency with academic language and justification.

KEYWORDS: academic language, collaboration, mathematics education, professional development, urban education

In the current high-stakes testing climate, instruction in many urban public school settings is becoming increasingly controlled and, in some places, scripted, as basic skills are prioritized over higher levels of reasoning. This narrowing of urban students' intellectual diet ultimately increases the education gap between these students and their more affluent counterparts (Anyon, 1997; Keiser, 2005). This education gap should not only be measured in mere test scores but also in the opportunities students have to learn to think and express themselves

MEGAN E. STAPLES is an assistant professor in the Department of Curriculum and Instruction in the Neag School of Education at the University of Connecticut, 249 Glenbrook Road, Unit 2033, Storrs, CT 06269; email: megan.staples@uconn.edu. Her research focuses on teaching and learning interactions in secondary mathematics classrooms, with a particular interest in collaborative interactions and detracked classrooms.

MARY P. TRUXAW is an assistant professor in the Department of Curriculum and Instruction in the Neag School of Education at the University of Connecticut, 249 Glenbrook Road, Unit 2033, Storrs, CT 06269; email: mary.truxaw@uconn.edu. Her research focuses on discourse in mathematics classrooms and teacher education, with particular interest in urban, linguistically diverse schools.

mathematically and reason in ways that will support their participation in a democracy and further their individual pursuits (Goodlad, 1994; Michelli, 2005).

In this article, we report the results from the *Mathematics Learning Discourse* (MLD) project, a small-scale research and development project undertaken with four teachers in two schools in an urban school district in Connecticut. The purpose of this project was to support teachers in fostering a *mathematics learning discourse* in their urban classrooms. Specifically, teachers sought to create a teaching and learning environment that developed students' academic language; promoted justification and argumentation (e.g., sense making); and provided all their students access to participation in cognitively challenging mathematical activities. Such an approach runs counter to typical pedagogy in urban settings (Leonard & Evans, 2008; Manouchehri, 2004). It was expected that this approach would enhance students' engagement and mathematics learning (Boaler & Staples, 2008; Brenner, 1998; Hiebert, Carpenter, Fennema, Fuson, Wearne, Murray, Oliver, & Human, 1997; National Research Council, 2001; Silver & Stein, 1996; Stein, Grove, & Henningsen, 1996; Wood, Williams, & McNeal, 2006), which, in turn, would increase their proficiency at responding to open-ended prompts, akin to those on the state standardized tests that often require higher levels of reasoning. We first discuss the research base for the MLD program and describe its implementation. We then evaluate the effect of the MLD project, analyzing student scores on pre- and post-assessments both for mathematical performance and for the development of students' proficiency with academic language and justification. We conclude with a discussion of the promise of the program's model and future next steps.

Focusing on a Mathematics Learning Discourse

We made several deliberate choices in developing the MLD program. One choice was to focus on student discourse. This focus is appropriate for two reasons. First, language is the predominant medium by which students learn and demonstrate their understandings. Language mediates learning (Vygotsky, 2002). Verbal discourse in classrooms (supported by symbolic representations, visuals, hands-on materials, etc.) is used to introduce students to mathematical ideas and provide opportunities to make sense of these ideas. How a teacher organizes her or his instruction to provide students access to, and opportunities for, meaning making and concept development is crucial to student learning. Acknowledging the centrality of language, we take the stance that it is through participation in practices such as justification and argumentation that students might expand their mathematical knowledge.

Second, the focus on student discourse is appropriate because it is not uncommon in urban settings to have students who are at various levels of language proficiency including English language learners (ELLs). In the school district of focus, nearly half of the public school students spoke a language other than English at home (Connecticut State Department of Education, 2008). As part of a needs assessment done prior to the project, teachers who were interviewed remarked that language issues affected their students' performance on state-mandated mathematics assessments. As noted in the literature, many ELL students have mastered conversational English, but have little exposure to academic language (Cummins, 2008). Thus, they face the double challenge of mathematics and language as they work on open-ended prompts. The development of students' academic language is a central function of schooling (Schleppegrell, 2007; Zwiers, 2008). Consequently, specific instructional practices within these schools, and other urban schools with similar characteristics, should support ELLs and bilingual students (Dalton & Sison, 1995).

The MLD Project: Research Basis and Rationale

The Mathematics Learning Discourse (MLD) project was undertaken during the 2007–2008 school year. In a 3-day summer workshop, a group of teachers from two urban public schools was introduced to the idea of a mathematics learning discourse. We presented this idea in terms of a model with three pillars (see Figure 1)—a model developed through a review of relevant research literature. We explored each pillar with the teachers through a series of activities and discussions. Given the research literature, it was expected that teachers who organize classrooms characterized by the three pillars might prove to be more effective with their students.

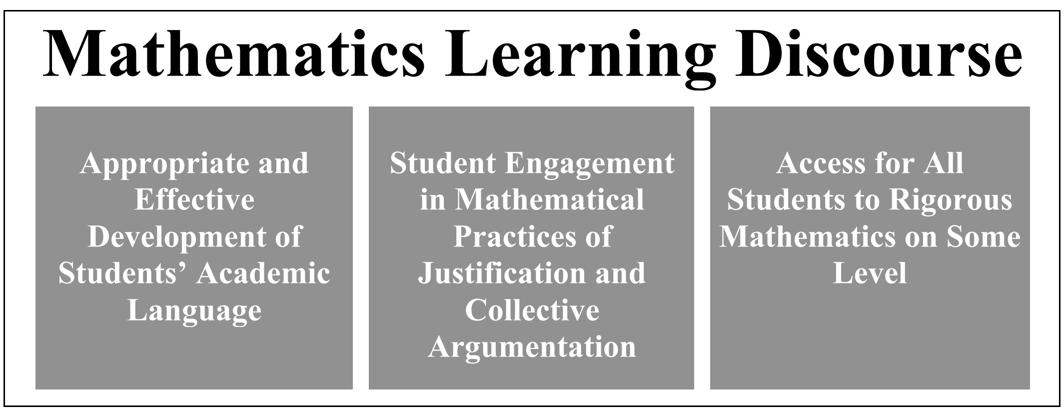

Figure 1. Three pillars of mathematics learning discourse.

Appropriate and effective development of students' academic language. A key aspect of developing students' academic language is to create a bridge from everyday, informal language toward academic language and use of the mathematics register (Halliday, 1978; Pimm, 1987; Zwiers, 2008). Too often, attention to language in mathematics classrooms focuses almost solely on vocabulary. Language-related instruction should move beyond simple vocabulary; it should include attention to how language is used to express mathematical ideas (functional linguistics) and the development of the mathematics register (i.e., language associated with the meanings of mathematics) (Halliday, 1978; Moschkovich, 2002; Pimm, 1987; Schleppegrell, 2007). This goal is pertinent for all students, but particularly so for students whose first language is not English (Cummins, 2000; Schleppegrell, 2007; Valdés, Bunch, Snow, Lee, & Matos, 2005). These students may be socially fluent, yet may need strategic linguistic support for engaging cognitively challenging mathematical tasks (Janzen, 2008), justification and higher order thinking.

Many classrooms, however, do not support such practices. Teachers remain unaware of the language demands involved in learning mathematics, especially as they pertain to justification and higher order thinking (Adler, 1999; Moschkovich, 2002; Pimm, 1987; Valdés et al., 2005). The effectiveness of attending to language development has been demonstrated by many researchers, including those who developed and implemented the SIOP® Model[1] (Echevarría, Vogt, & Short, 2007, 2010), which became one central feature of working on academic language with this group of teachers. We drew on SIOP® strategies and techniques because they have been proven effective for supporting the teaching of academic content to ELLs (Echevarría, Short, & Powers, 2006).

Student engagement in mathematical practices of justification and collective argumentation. Student participation in justification, meaning making, and argumentation have been implicated as critical components for supporting students in learning mathematics (see, e.g., Hiebert et al., 1997; National Research Council, 2001; Silver & Stein, 1996; Stein, et al., 1996; Wood et al., 2006). There is also some evidence that participation in these practices is particularly effective in supporting the learning of lower attaining students and/or ELLs (Boaler & Staples, 2008; Moschkovich, 2002). By justification and argumentation, we mean engaging students in the processes of sense making (Hiebert et al., 1997) and having them offer claims, supported by evidence and warrants (Toulmin, 1958) in order to support a result and convince others of the claim's validity.

Access for all students to rigorous mathematics on some level. In addition, teachers need to ensure that all their students, who vary in their prior mathematical background, language ability, and other characteristics, have access

[1] SIOP®, formerly known as the Sheltered Instruction Observation Protocol, is an instructional approach that offers teachers a framework for planning and implementing high quality instruction for English language learners.

to participate in the lesson's mathematical activities (Goodlad, 1994; Michelli, 2005). Inequitable access and participation to mathematics during class leads to inequitable learning opportunities and learning gains (Cohen, Lotan, & Leechor, 1989; Cohen & Lotan, 1997; Gee, 2003; Martin, 2003). Components of access for all were conceptualized to include developing productive classroom norms for discussion (Yackel & Cobb, 1996; Wood, 1999) and groupwork (Cohen, 1994a, 1994b), using manipulatives, and designing tasks that allow a range of students access to engaging the task. These tasks were multi-dimensional (Cohen, 1994a; Lotan, 2003) and cognitively demanding (Stein et al., 1996). In addition, *access for all* attended to explicitly teaching what a good justification looks like and providing formative feedback (Black & Wiliam, 1998) to support student learning.

The MLD Project: Development and Practices

In the summer of 2007, a group of teachers from an urban school district participated in 3 days of summer professional development (PD) that focused on the three pillars. Teachers participated in a range of activities and discussions related to the following: strategies to support language development, with special attention to ELL students (Echevarría, Vogt, & Short, 2007, 2010); analysis of cognitively challenging tasks (Stein, Smith, Henningsen, & Silver, 2000), including language demands related both to making sense of the task and offering justifications; and strategies to support engagement and access for all students (e.g., strategic groupwork, formative feedback, etc.).

For example, we analyzed the cognitive demands of tasks by adapting an activity from Stein et al.'s (2000) *Implementing Standards-based Mathematics Instruction,* and discussed various task features that could be modified to ramp up the cognitive demands (contrasted with those that made a task more complicated or harder to access). Before this activity, we discussed what higher order thinking meant—a discussion we revisited throughout the summer sessions and academic year.

Focusing more on language, a colleague in bilingual education worked with the group to introduce them to the elements of the SIOP® Model—for example, explicit inclusion of language objectives (along with content objectives) within mathematics lesson plans. Additionally, she guided the group in analyzing the linguistic demands of open-ended prompts from our state assessments. We also introduced and modeled strategies such as math *talk moves* (Chapin, O'Connor, & Anderson, 2003) and analyzed and discussed transcripts and videos from mathematics classrooms (Boaler & Humphreys, 2005) considering how the classroom discourse may promote conceptual understanding and higher order thinking.

Additionally, teachers participated in and reflected on cooperative problem solving activities that were strategically designed and implemented to enhance access for all in mathematics classrooms. These activities pressed for higher order thinking and included the use of "check points"—points in the task where the teacher is called over and all members of the group need to be prepared to explain the work and respond to questions. These check points provided an opportunity for formative assessment (as the teacher interacted with the group) as well as language development.

The summer experience was followed by ongoing collaborative work across the course of an academic year involving the development, implementation, and debriefing of *higher order thinking* (HOT) mathematics lessons. The HOT lesson plans incorporated pedagogical strategies related to each of the three pillars (e.g., content and language objectives, verbal and written discourse that provided opportunities for higher order thinking and justification, and strategic support to allow all learners to engage in meaningful mathematics). The collaborative teams included teachers, university teacher educators/researchers, and preservice mathematics teachers who completed internships in the schools. The collaborative meetings took place weekly (planning one week and debriefing the next), and the HOT lessons occurred approximately twice each month lasting about 1 hour each. The HOT lesson plans were archived for public use (see the following website for archived lessons: http://www.crme.uconn.edu/lessons/). Along with developing, implementing, and reflecting on HOT mathematics lessons, the teachers were encouraged to infuse the three pillars of mathematics learning discourse in their everyday teaching practices. (For examples of practices associated with each of the pillars, see Truxaw & Staples, 2010.)

The MLD Project: Research

To document the possible impact of the project and evaluate the potential of the underlying model, we address the following research questions:

1. What was the impact (if any) of the Mathematics Learning Discourse project on student performance on open-ended math prompts?
2. What was the impact (if any) of the Mathematics Learning Discourse project on student demonstrated proficiency with academic language and mathematical justification?

In examining these questions, we focus primarily on student learning data. These research questions allow us to address a broader question of interest: *Does the MLD model seem to hold promise as an approach to professional development in urban schools?* In referencing the MLD model, we intend to indicate both the conceptual underpinnings for the "content" of the work (the three pillars), and the program's structural design (summer sessions with yearlong follow up in collaborative teams). We first report on findings related to the two research questions.

We take up the question of the model's promise in the *Discussion and Implications* section.

The project took place in five classrooms of four teachers in two urban schools (one K–8 school and one high school) in Connecticut—one grade 4, one grade 5, and three grade 9 classes (one teacher had two sections of algebra and one had one section). The schools had partnered with the researchers' university on other projects so there was already some level of rapport established. The grade 4 and 5 teachers had 18 and 29 years of teaching experience, respectively, and the grade 9 teachers had two and three years of teaching experience. The teachers volunteered to be involved with the project and received a small stipend for their participation.

In the focus school district, 95% of students qualified for free or reduced-priced meals and 94% of the students were categorized as "minority" students. Additionally, at the two schools more than half of the students spoke a language other than English at home (52% at the K–8 school and 71% at the high school) (Connecticut State Department of Education, 2008).

Data Sources. The principal source of evidence to address the research questions examined in this article was student pre- and post-assessment performance data on open-ended mathematics prompts from participating teachers' classrooms as well as other classes within each school. For each of the grade levels, the prompt was a released or sample item from the state tests that required higher order thinking and/or a justification of the student's response. The state assessments include two open-ended prompts for grades 3–8 and four open-ended prompts for grade 10. Students were given up to 30 minutes to complete the pre- and post-assessment prompts. All prompts were embedded in some context (as this is the priority of the state). We also administered a reflective survey after students completed the assessments. We asked students to restate the problem in their own words, circle confusing words and phrases, and identify other aspects of the problem that were confusing or difficult for them.

To gauge whether changes in students' performance in these classes were beyond what could be expected in a typical year, we collected data from the same single-item pre- and post-assessment from other classes at these schools. For grade 9, we collected data from eight other ninth-grade classes. For grades 4 and 5, we administered the same prompts in the project teachers' classes the year prior to their involvement with the project.

We chose to use prompts from the state assessment for several reasons; most notable of these was the tremendous pressure on teachers to improve student performance on such assessment. The teachers also identified these prompts as challenges for their students. For instance, in 2008, 40% of 8th graders across the state achieved "mastery" on this component of the state tests; in the urban school district of focus, only 12% of 8th graders achieved mastery. In addition, these prompts are language intensive (generally both reading and writing are necessary) and require higher order thinking. Thus, they aligned with the goals of the project.

Although the analysis for this article focuses on the student performance data, other data collected for the project include: materials and field notes from the summer PD; HOT lesson plans; student work samples from HOT lessons; audio-recordings and field notes of lesson planning and debriefing sessions as well as the implementation of HOT lessons; and teacher interviews.

Data Analysis. We analyzed the pre- and post-assessments from multiple perspectives. Prompts were scored using the state rubrics for open-ended prompts; two trained scorers independently scored each prompt. If scores differed, another scorer scored the prompt. The scores were analyzed using descriptive statistics, as will be presented in the *Findings* section. The analyses we conducted varied by grade level depending on the data available. For the ninth-grade classes (two teachers), we made two main comparisons. First, we considered all students who completed the prompt and compared the scores of students in MLD classes with those in non-MLD classes. Second, we considered only students who were in MLD classes all year and compared their results with those students who were in non-MLD classes all year. This reduced our sample size, but may provide a more accurate picture of the possible effect of the project. For grades 4 and 5, we did not have a large comparison group. Rather, we compared end-of-year scores on identical prompts for the MLD teachers' classes from the year prior to the project and MLD teachers' classes for the project year, allowing for group-level comparison between the two classes.

To directly target student academic language and justification, we developed a rubric to score student work samples by applying research literature related to argumentation and justification (e.g., Healy & Hoyles, 2000; Toulmin, 1958) and academic language/mathematics register (e.g., Pimm, 1987; Schleppegrell, 2007). The initial rubric described different levels of proficiency with respect to two categories: use of academic language and argument/justification. For argument/justification, we used as a working definition: the process of sense making (Hiebert et al., 1997) to remove doubt about a claim using logical reasoning, including evidence of claims, warrants and evidence (Toulmin, 1958). For academic language, we used the working definition: language appropriate to communicate the mathematics involved in the context of the problem/situation, including processes, properties, functions and relations (Halliday, 1978; Pimm, 1987). Academic language entails use of vocabulary that is important for expressing ideas precisely and mathematically, as well as use of appropriate sentence structures, and so forth, that are needed to express mathematical ideas (e.g., generalizations, justifications, identification of a counterexample, etc.).

We made some adjustments to the rubric as we applied it to student work samples. For example, we realized that the overlap of justification and academic language was quite extensive on many prompts, as expressing a justification relied upon using language to express causal relationships and inferences. This overlap was particularly extensive with contextual problems that required little in terms of specific mathematical terminology; therefore, most of the academic language required to respond to the prompt was related to expressing the justification. We modified the rubric (see Table 1) to indicate this overlap by showing a middle band that we could not attribute to either academic language or justification alone. The use of specific mathematical terminology (e.g., product, polygon) was attributed exclusively to the Use of Academic Language category. The types of inferences and soundness of the reasoning students used was attributed exclusively to the Argument/Justification category.

Table 1
Academic Language and Justification (ALJ) Rubric

Score	**3**	**2**	**1**	**0**
Use of Academic Language	Student uses appropriate mathematical terminology consistently (e.g., "equals" *vs.* "makes")	Student uses appropriate mathematical terminology, as required by prompt.	Few uses of mathematical terminology are present, as required by prompt.	Student work reveals little or no command of academic language or use of mathematical terminology.
Use of Academic Language and Argument/ Justification	Student work reveals appropriate words and/or phrases to indicate logical connections and relationships; claim is expressed for all relevant cases.	Work reveals some appropriate indicators of logical connection and/or relationships; claim may or may not be expressed for all relevant cases.	Work demonstrates challenges with articulating logical connection and/or relationships; claim is not expressed for all relevant cases.	Work reveals little or no language that describes requisite logical connection or relationships; claim is not expressed, or not expressed for all relevant cases.
Argument/ Justification	Claim holds for all relevant cases; argument demonstrates validity for all relevant cases; student offers a justification that includes a claim and explicitly identifies evidence, as well as the logical connection between the claim and evidence.	Student offers a justification that includes a claim and/or a warrant and/or evidence; connection between the claim and evidence is partially articulated; justification may or may not hold for all required cases.	Student offers a claim and may have work that supports the claim, but the student does not make the connection between these explicit.	Student does not produce work that includes a justification or claim, or student only offers a minimal response as a claim that could be seen as a guess.

For each task, we operationalized the rubric based on the demands of the task and nature of required justification. Refinements were made as we considered student work samples, and then rescoring was done using the refined criteria. Based on the demands of the particular prompts we used, we opted to give a single score (rather than separate scores for each category). Thus we scored holistically but accounted for all three categories of the rubric, as necessitated by the task. Twenty percent of prompts were double scored to ensure consistency across scorers, and one scorer scored the remainder.

Limitations. Although the results we share are promising, we raise a caution to the reader that these results suggest value in this model and approach, but are not definitive. More work needs to be done. The results reported here are based primarily on a single prompt, albeit open-ended and requiring extensive thought and one that is drawn from the state assessment program and consequential for students' performance on state tests. Any one prompt may have unique features that may unknowingly impact the particular results. For example, Solano-Flores and Trumbull (2003) found that, for ELLs, "each [assessment] item poses a different set of linguistic challenges" and that "ELL performance varies considerably not only across items but also across languages" (p. 8). Similar results may also hold for non-ELL students who are developing their academic language and proficiency with justification.

With respect to gauging students' proficiency with academic language and with justification, it is important to note that content understandings are a confounding factor for any score related to justification or academic language (just as academic language is a confounding factor for students' demonstrated proficiency with mathematical content on any question that requests a justification). Without some level of content understandings and ability to read and comprehend, it is impossible for a student to demonstrate her or his level of proficiency with academic language or justification on these contextual prompts. Thus, if a student leaves a prompt blank, it may be that the student does not understand the material or that the student could not access the problem (reading comprehension) and determine the mathematical work required by the prompt. This complexity is near impossible to sort, and is one reason why it is so challenging for teachers to focus on student growth in these areas, which we discuss later.

Findings

In this section, we present evidence of student improvement in demonstrated proficiency regarding both content and academic language/justification, and subsequently argue for the value of this model for professional development to support students' engagement in higher order thinking and the development of their academic language.

Student Mathematical Performance

Prompt score gains. Students in participating classrooms in grades 4, 5, and 9 demonstrated improved performance on open-ended prompt scores. A score of 2 or 3 was considered in the "mastery" range (a term used by the state). Across the

MLD classes, the level of mastery increased overall. Additionally, there were marked decreases in scores of 0 across all three grade levels. We review the results for each grade level and offer evidence that the improvement was greater than would be expected without the professional development program.

Grade 9 Results – State Scoring Rubric

The following prompt was administered in the ninth-grade classes, which included the three MLD classes and eight other ninth-grade classes for purposes of comparison:

> For an original graphic design, Lee charges a fixed fee of $50 plus $25 for each hour that he works. His main competitor charges a fixed fee of $40 plus $30 for each hour that he works on a design. Lee's competitor advertises that his rates are cheaper. Is Lee's competitor correct? Explain your reasoning. (The grid is provided in case you decide to use a graph as part of your explanation.) Remember to show your work. (2003 released item, Connecticut State Department of Education, 2009)

This prompt requires students to understand the fee structure for Lee and his competitor and to determine a way to assess the validity of the competitor's claim. A full analysis reveals that Lee is cheaper for any job that takes longer than 2 hours; his competitor is cheaper for a job that takes less than 2 hours; and they charge the same amount at 2 hours. Students can solve this problem in a wide variety of ways, including graphing, using equations, generating specific points, and analyzing the relative rates of change. We observed all strategies being used.

Table 2 reports the results of the fall administration (pre-assessment). These data show that the students across the ninth-grade were not faring well on this prompt and that students in the ninth-grade MLD classes were generally doing more poorly. Eighty-eight percent of the students scored a 0; only 4% were considered at mastery level.

Table 2

Ninth-Grade Student Scores on Open-ended Prompt, Fall (Pre-Assessment)

	Number of Students			**Score**		**Percent Mastery**
	Fall	0s	1s	2s	3s	
MLD Classes	49	43 (88%)	4 (8%)	0 (0%)	2 (4%)	4%
Non-MLD Classes	131	99 (76%)	14 (11%)	10 (8%)	8 (6%)	14%
All Ninth-Grade Classes	180	142 (79%)	18 (10%)	10 (6%)	10 (6%)	12%

Table 3 reports the post-assessment scores showing the improvement of the MLD classes and the relative steadiness of the results from the other ninth-grade classes. Whereas the number of 0s in the MLD classes decreased dramatically from 88% to 47% and the percent mastery increased from 4% to 31%, the distribution of scores of other ninth-grade students remained approximately the same between fall and spring. Note that the attrition/retention rates for the MLD classes are similar to those of the larger sample of ninth-grade students.

Table 3

Ninth-Grade Student Scores on Open-ended Prompt, Spring (Post-Assessment)

	Number of Students	Score				Percent mastery
	Spring	0s	1s	2s	3s	
MLD Classes	38	18 (47%)	8 (21%)	5 (13%)	7 (18%)	31%
Non-MLD classes	100	71 (71%)	14 (14%)	10 (10%)	5 (5%)	15%
All Ninth-Grade Classes	138	89 (64%)	22 (16%)	15 (11%)	12 (9%)	20%

We also considered the scores of the subset of students who completed the prompt in both the spring and the fall (30 students in the MLD classes and 76 students in non-MLD classes, for a total of 106 students). The data showed the same trends held when considering only that subset of students, although the overall results of this subset of students were slightly better, perhaps not surprisingly. Overall, these are positive results for the ninth-grade MLD classes.

As we conducted our scoring, we looked for evidence of growth of academic language and student justification that might be captured by using the state rubric. We found two indicators that students in the MLD classes were outperforming those in the non-MLD classes with respect to justification and academic language.

First, students in the MLD classes were more likely to maintain a connection between the mathematics they were doing and the context of the prompt, and to directly relate their mathematical work to making an argument about the competitor's claim. For example, in the spring, there were similar percentages of students in the MLD and non-MLD classes using algebra, 33% and 35%, respectively. This method is relatively sophisticated, requiring students to write two functions modeling the two different price schemes. Students can then generate a table of values, or can solve the system of equations (to find where the costs are equal). To successfully answer the prompt question, students must make a connection between the solution of the system and its meaning in context of the problem. Many non-MLD students offered no (or an erroneous) interpretation of the meaning of the results of their solution method. They simply wrote $x = 2$ (the correct result obtained when solving the system of equations). In the MLD classes, more students connected their mathematical work with the context and offered an interpretation of their results. Of students using algebra, 46% of the MLD students received a score of 2 or 3, compared to 24% of those in the non-MLD classes, a score that could only be obtained with some interpretation of the results in the context of the problem.

The second indicator was whether a student attempted a written response. A blank response is often a sign of being overwhelmed or not being able to comprehend the word problem. Thus, a shift from a blank paper (or the response "I can't do this") to some writing likely indicates increased comprehension of the problem. We conducted a simple count on the number of students who left the problem blank (no writing and no calculations). The percentage of students leaving the problem blank decreased for both MLD and non-MLD classes, but considerably more for the MLD classes. For the non-MLD classes, 21% left the problem blank initially and 18% left the problem blank at the end of the year. For the MLD classes, the initial percentage was 29%, and this was reduced to 10% at the end of the year. For the 30 students in the MLD classes who completed both the spring and fall prompt, 37% (11 of 30) offered some written explanation in the fall and 70% (21 of 30) offered some written explanation in the spring, nearly double the fall numbers.

Although it is difficult to determine whether we should identify this result as evidence of increased proficiency with academic language, we take it as a potential positive indicator of language use.

Grade 9 Results – Academic Language and Justification Scores

To explore changes in students' proficiency with academic language and justification in a more targeted way, we applied the Academic Language and Justification (ALJ) rubric (Table 1) to the Graphic Design prompt. We scored the prompts of the 30 MLD students for whom we had both pre- and post-scores. One important challenge that emerged related to tracking students' academic language was the nature of the Graphic Design problem itself. This particular prompt required academic language to structure a justification appropriately (stating claims, warrant and evidence and linking those) and to use language of comparison (which strongly overlaps with everyday language). There was little need, however, for students to use mathematical vocabulary (an aspect of academic language), perhaps in part because the prompt is couched in a "real-world" context. Thus, our scoring was based primarily on two of the three components—the one that indicates the overlap of justification and academic language, and the justification category. For example, a student could write, "I think Lee's competitor is wrong. For 5 hours, Lee's competitor charges more. See my table below."

Mathematically, this is a complete justification. However, such a justification does not require much use of mathematical academic vocabulary. As a second example, a student might write, "My table shows that Lee is more expensive for 1 hour. They charge the same price for 2 hours. Lee is cheaper for more than 2 hours. So Lee's competitor is not correct."

We would like to highlight one difference with the state rubric that we found particularly interesting. The prompt asks whether the competitor's claim is correct. From a mathematical standpoint, as well as legal, the student needs to find only one counterexample to Lee's competitor's claim that his (the competitor's) rates are cheaper to prove the statement false. A few students in the spring indeed found one counterexample (e.g., at 3 hours) and claimed Lee's competitor was incorrect. Our interpretation of the state scoring rubric, however, indicates that the assessors would award 1 out of 3 points for this response. The point would be awarded for showing appropriate calculations for one data point. Although this argument structure (specifically, showing one counterexample to prove a statement false) is mathematically sound, and potentially reflects a sophisticated understanding of the problem and role of a counterexample, the state required students to offer a full analysis of the situation (including identifying the "break-even" point) to receive a score of 3. The problem, however, does not explicitly indicate this requirement. Thus, we felt that some students who had a sophisticated mathematical understanding may have scored low on the state prompt. The justification rubric, however, counts this strategy as a valid approach to justifying one's response that the competitor is not correct.

Table 4 reports sample student responses at each of the scoring levels, along with a commentary that indicates some features noted for determining the score.

Table 4

Scoring Examples for the Graphic Design Academic Language and Justification Rubric (Grade 9)

Score	Examples	Commentary
0	a. Left blank. b. Shows only an equation for each cost structure. c. "Lee's competitor is right"; no other work is shown.	These responses provide no indication that the student formulated an argument or could express it.
1	"The competitor is right." Shows computations of 40+30 = 70, labeled *Competitor*. Shows computations of 50+25 = 75, labeled *Lee*.	Offers claim and evidence. Evidence is not linked to claim; warrant is implicit. No use of academic language such as "because."

2	"His rates would be even, it depends on each hour. So each one is not cheaper neither the highest amount. They both are the same amount counting each hour." $50 + $25 + $25 = $100 $40 + $30 + $30 = $100	Claim is implicit (the competitor is not correct); includes a warrant (the rates can be even) and offers evidence (showing computations) linking them by saying "they both are the same amount." Includes some indication that who is cheaper varies by the hours worked ("it depends").
3	"No Lee's competitor is not correct because if it takes him three hours or more the price is more expensive than Lee's." Work shows table for hours and dollars for Lee and competitor, with all values labeled.	Offers a claim, a warrant (there are jobs for which the competitor is more), and points to the evidence ("if it takes him 3 hours or more" and includes a table). Logical connectors used (because) and evidence explicitly linked. The subordinate clause (if it takes…) specifies the domain for which Lee's competitor costs more (3 hours or more).

Table 5 reports the students' score results using the ALJ rubric for the 30 students in the MLD classes who completed both fall and spring prompts. In this group of students, we see a clear trend towards higher levels of demonstrated proficiency with academic language and justification. In the spring, students were more successful in constructing a valid and complete argument and in using appropriate language to present their argument and response to the question. As noted, content understandings are required to demonstrate proficiency in these areas; therefore, some of the increase in these scores is likely a product of improved content understandings. Similarly, changes in scores could reflect an increase in reading comprehension and understanding particular language such as "fixed fee," which surveys revealed many students did not understand. We cannot identify all the factors that contributed to this change. We certainly expect an influence by the work the MLD teachers were doing with their students. It could also be, however, that students' exposure to other learning opportunities, such as an outstanding English course, had a considerable effect.

Table 5
Ninth-Grade Student Academic Language and Justification Scores on Open-ended Prompt (Pre- and Post-Assessment, *n* = 30)

	Score			
	0s	1s	2s	3s
MLD Classes Pre-Assessment	17 (57%)	8 (27%)	3 (10%)	2 (7%)
MLD Classes Post-Assessment	9 (30%)	6 (20%)	5 (17%)	10 (33%)

To see if we might more specifically identify and describe the nature of changes between students' fall and spring prompts, we also examined each pair of student responses side by side. For the most part, we found that we were unable to make definitive judgments about whether differences noted in a student's responses between the fall and the spring offered specific evidence of a change in their proficiency with the practice of justifying, their use of academic language, or their mathematical understandings. These areas are naturally interdependent and our data were not sufficient to allow us to parse these components for all students. There were some instances, however, where, due to the particular nature of the responses, we felt that we could make a claim about one of these three areas. We offer some examples and analyze the nature of the changes we observed.

In the following example, we argue that we can see growth in a student's use of academic language in producing a justification. Below is the written component of the student's fall and spring responses:

> **Fall:** His rate is cheaper cause if you add the Fee and 1 hour pay his charge of money is larger.
> **Spring:** So Lee's competitor is not cheaper. It cheaper for the first hour. The second they are even. But the next hours Lee is cheaper as you see on my graph…. (Student 11103)

In terms of language, we see that the first response does not explicitly name Lee or the competitor. Each is referenced once by the term "his." Thus, there is an increase in the specificity of references. The first response also does not offer a claim that directly addresses the question "Is Lee's competitor correct?" whereas the second response does in the first sentence. Mathematically, we see growth as well. In the fall, the student only considered what happened at 1 hour, despite seeming to understand (by the written response and her work) that there was an hourly fee. In the spring, the student analyzed the situation, fully considering all possible cases for the number of hours.

In the second example, we see a difference in a student's mathematics, which provides more evidence of the student's academic language use. We are not certain, however, if there is evidence of growth in both of these areas, or just mathematics (which then provided us the opportunity to see more academic language). One interesting feature of the following example is that both the fall and spring responses have the same structure, namely, the student makes a claim, uses the linking word *because*, and then provides evidence. Thus, we do not see a difference in how the student produces a justification.

> **Fall:** No Lee's competitor is not correct because Lee's competitor is cheaper. (The student's calculation demonstrates computations for a 1-hour job for each.)
> **Spring:** Lee's competitor is not correct because when I did a line graph Lee's competitor line increased more than Lee's. (The student has a graph for Lee and his competitor, demonstrating costs for hours 1 through 5.) (Student 11109)

Note that, mathematically, the first response is incorrect. The evidence does not support the claim. Furthermore, the student only explored a 1-hour job. In the spring, however, the student produced a set of values and a graph, referencing the graph as well as a particular feature of the graph (the line increased more) as evidence to support his claim. Through his academic language, we could "hear" him interpret his graph using mathematical language.

Grade 4 Results – State Scoring Rubric

To gauge impact on student performance in grade 4, we report on scores from identical fourth-grade prompts that were administered to the MLD teacher's non-MLD fourth-grade class prior to the MLD project (spring 2007) and to the MLD fourth-grade class near the end of the project year (spring 2008). This comparison allows for a group-level comparison between the two classes. The prompt is shown in Figure 2.

Hot Dog Buns

You estimate that you'll need 40 buns for a class picnic.
Hot dog buns are sold in packages of 8 and packages of 12.

- The package of 8 costs $1.00
- The package of 12 costs $1.20

a. Show three different ways you could buy packages to get at least 40 buns.
b. Which packages would you buy if you wanted to spend the least money? Show or explain how you arrived at your answer.
c. Which packages would you buy if you wanted exactly 40 buns?

(Sample Item, Connecticut State Department of Education, 2009)

Figure 2. Fourth-grade prompt.

This prompt required that students work through multiple tasks and constraints, including mathematical, contextual, and linguistic challenges. The language challenges included some unfamiliarity with "everyday" words and phrases (e.g., *package* and *purchase*), as well as other words that are germane to the mathematical work they are expected to engage. For example, phrases such as "*at* least" and "*the* least" might seem familiar, but may be misinterpreted or not understood, leading to very different mathematical work. In terms of mathematical content, students need to demonstrate facility with estimating, display fluency with addition and multiplication of whole numbers and money amounts expressed as decimals, and be able to compare and evaluate different solutions.

To achieve a mastery score on the prompt (score of 2 or 3), according to the state rubric, students needed to provide appropriate answers for at least 2 of the 3 parts to the prompt. For part *a*, they could show pictures, numbers, words, or diagrams to show the different ways to make at least 40 buns. For part *b*, it was reasonable to make comparisons from responses to part *a*. Although the least cost overall is $4.40 (2 packs of 12 buns + 2 packs of 8 buns →(2 x $1.20) + (2 x $1.00) = $4.40), given the grade level, it was considered acceptable if students compared from the three combinations made in part *a*, selecting the least of the three, or if they articulated a reasonable cost per item explanation. However, it was not acceptable if students simply stated that the packs of 8 buns cost less because $1.00 is less than $1.20 (the cost of 12 buns). For part *c*, the students needed to show a combination equal to 40 buns exactly (e.g., 5 packages of 8 buns = 40 buns). This prompt was selected because it was typical of those used in the state assessment; it required interpretation of everyday and academic language, and, although it did not strictly require them, it provided opportunities for written explanation and justification.

Table 6 reports the results of the administration of the prompt using the state scoring rubric for the classes of our participating fourth-grade teacher in the year prior to the project (spring 2007) and for the project year (spring 2008). Although the two classes of students differed, the prompts were identical, both classes were heterogeneously grouped, and both classes were taught by the same teacher. For the MLD class, scores were considered only for students who were in the class the full year. The spring 2008 (MLD) class demonstrated greater mastery (43%) than the same teacher's class prior to MLD involvement (22%)—nearly double the rate of mastery of the students from the previous year. Further, the MLD class showed a lower percentage of scores of 0 (14% MLD vs. 56% pre-project). The results suggest an impact of the MLD project on student performance on open-ended prompts.

Table 6

Fourth-Grade Student Scores on Open-ended Prompt, Spring (Post-Assessment)

	Number of Students	**Score**				**Percent Mastery**
		0s	1s	2s	3s	
Pre-MLD Class (Spring 2007)	18	10 (56%)	4 (22%)	3 (17%)	1 (6%)	22%
MLD Class (Spring 2008)	21	3 (14%)	9 (43%)	5 (24%)	4 (19%)	43%

Note: Rounded to closest whole %; therefore, may not sum to exactly 100%.

Grade 4 Results – Academic Language and Justification Scores

To explore growth in grade 4 students' proficiency with academic language and justification, we adapted the general ALJ rubric for use with this grade level and for the specific prompt. The spring Hot Dog Bun prompt had three parts (*a*, *b*, *c*). Because part *b* ("Which packages would you buy if you wanted to spend the least money? Show or explain how you arrived at your answer.") included the greatest opportunity for academic language and justification, it was the main focus of the ALJ scoring. Indicators for the ALJ rubric were identified that related to this part of the prompt in particular, but that also considered the other two sections. For example, to justify part *b*, students typically needed to refer to work done in part *a* ("Show three different ways you could buy packages to get at least 40 buns.") and make an argument for which of their three ways was the least expensive. Table 7 shows sample responses at each of the scoring levels, along with some commentary.

Table 7
Scoring Examples for Academic Language and Justification Rubric Hot Dog Bun Prompt (Grade 4)

Score	Examples	Commentary
0	"I would buy the low price ones. This was my answer because I don't want to waste my money."	Little command of academic language; claim does not have a logical connection to question being asked; no relevant evidence or warrants.
1	"I would take the package of 8 because it cost \$1.00 and \$1.00 is less than \$1.20."	Includes a claim (minimal use of academic language), a "because" statement, but does not include logical connection to the context of the problem.
2	Student shows computation of prices for two different combinations of packages. Show prices \$4.40 and \$5.00 and says, "easy compare" and shows the claim, "2 of the 8 dog packs and 2 of the twelve" (as the least expensive).	Uses language—e.g., packs, compare, twelve. (In parts *a* and *c*, used "packages of.") Claim is shown ("2 of the 8 dog packs and 2 of the twelve"), though not clearly labeled as "least expensive." Evidence (student work) and warrants ("easy compare") are shown. The reader needs to infer logical connections.
3	Student refers to own work from part *a*, organized as 1, 2, and 3, presenting the claim, "My third idea" followed by "because it is \$4.40 and my 2 and 1 idea are a higher price." The student then lists the three prices, and circles the lowest of the three.	Demonstrates ability to use academic and appropriate contextual language to represent responses. Claim is explicit ("My third idea" [is least expensive]), provides a warrant "because it is \$4.40 and my 2 and 1 idea are a higher price" (compares with prices for other two cases), and evidence (refers to work showing prices for each case).

Table 8 reports the ALJ scores for the fourth-grade classes of our participating fourth-grade teacher in the year prior to the project (spring 2007) and in the project year (spring 2008). The MLD class (those who participated the full year) demonstrated higher percentages of mastery, scores of 2 or 3 (28%), as compared with the class before MLD involvement (6%). Further, the MLD class demonstrated lower percentages of scores of 0 (14% MLD; 39% pre-MLD). These results suggest that, overall, the MLD students demonstrated greater proficiency with academic language and justification than students who did not participate in the project—on the same prompt, at the same time of year, with the same teacher.

Interestingly, there was not a clear relationship between the state scoring and ALJ scoring. For example, of the fourth-grade students who achieved mastery (scores of 2 or 3) according to the state rubric, only 60% scored a 2 or 3 using the ALJ rubric—suggesting that the state rubric does not attend directly to academic language and justification, even on open-ended prompts. Overall, approximately 23% of the students scored lower on the ALJ rubric than the state rubric, approximately 18% scored higher on the ALJ rubric than the state rubric. These results suggest that rubrics such as ours might be necessary to unpack students' proficiency with academic language and justification.

Table 8
Fourth-Grade Student Scores for Academic Language and Justification on Open-Ended Prompt

	Number of Students	**Score**			
		0s	1s	2s	3s
Pre-MLD Class (Spring 2007)	18	7 (39%)	10 (55%)	1 (6%)	0 (0%)
MLD Class (Spring 2008)	21	3 (14%)	12 (57%)	3 (14%)	3 (14%)

Note: Rounded to closest whole %; therefore, may not sum to exactly 100%.

Grade 4 Results –Student Perceptions of Prompt Challenges

As noted in the *Data Sources* section, students were asked to reflect in writing about the prompts. In grade 4, student reflections revealed some awareness of contextual and linguistic challenges. For example, responses related to what was confusing included: "I was get confused when they said if you wanted to spend the least amount of money," and, "another thing that was confusing to me is that the question was telling me stuff I couldn't understand." Of course, not all students were able to recognize and/or communicate what they did or did not know. For example, many fourth-grade students who did not achieve mastery indicated that they did not find anything confusing about the problem (e.g., "Nothing was difficult for me."). Others circled whole questions or the entire page. These are indicators that parsing out what was confusing seemed challenging for many of these students. This result is of concern given that students' facility with unpacking academic and contextual language in problems can influence their ability to make sense of what is required of them mathematically. This concern may be particularly pertinent for ELLs who are socially fluent but have not yet achieved academic fluency, as ELLs' social fluency may mask their need for linguistic support in academic contexts such as written prompts (Cummins, 2000).

Grade 5 Results – State Scoring Rubric

To gauge impact on student performance in grade 5, we report on scores from identical fifth-grade prompts that were administered to the MLD teacher's non-MLD fifth-grade class prior to the MLD project (spring 2007) and to the MLD fifth-grade class near the end of the project year (spring 2008), allowing for a group-level comparison between the two classes. The prompt is shown in Figure 3. This prompt was selected because it had similar demands to those described for the fourth-grade prompt; that is, it included the need to consider and work through multiple tasks and constraints that included mathematical, contextual, and linguistic challenges. In order to achieve a mastery score on the prompt, according to the state rubric, students needed to demonstrate reasonable estimates for the number of burgers and rolls based on the information given (approximately 85 hamburgers and 65 rolls), along with determining how many of each item is "enough" to meet the estimates (should be equal to or greater than estimates—within a reasonable range). Additionally, students needed to calculate the cost of the packages and total cost accurately based on the number of each type of package purchased. A limitation of the prompt in terms of the ALJ scoring was that, although it says to "show how you arrived at your answer," academic language was not required. Consequently, we were unable to apply the ALJ rubric to analyze the fifth-grade prompts. We report here only the scores from the state rubrics and some themes drawn from the students' written reflections.

Table 9 reports the results of the administration of the prompt using the state rubric for the classes of our participating fifth-grade teacher in the year before the project (spring 2007) and for the project year (spring 2008). Although the two classes of students differed, the prompts were identical and were taught by the same teacher. For the MLD class, scores were considered only for students who were in the class the full year. The spring 2007 class (pre-MLD) was ability grouped (high ability); the spring 2008 group (MLD) was heterogeneously grouped. This "ability" grouping would suggest the likelihood that the pre-MLD class' scores would surpass those of the MLD class. The mastery level of the MLD class (43%) that was not ability grouped was higher than the same teacher's class the prior year (33%) that was identified as a high ability group. Further, the MLD class showed a lower percentage of scores of 0 (13%) than the pre-MLD class (29%).

Hamburger Rolls

You are going to have a Fourth of July picnic for friends and family. You estimate that:

- 25 people will have 2 hamburgers and 2 rolls each;
- 15 people will have 1 hamburger and 1 roll each;
- 20 people will have 1 hamburger and no roll each.

Hamburgers and rolls are sold two ways each:

Hamburgers	Rolls
8 hamburgers for $1.75	6 rolls for $0.75
12 hamburgers for $2.15	18 rolls for $1.80

Use this information to order enough hamburgers and rolls for the people coming to your picnic. Show how many packages of each size of hamburgers and rolls you will buy. Compute the final cost of all the items. Show how you arrived at your answer.

Items	Number of Packages	Cost
8 hamburgers/$1.75		
12 hamburgers/$2.15		
6 rolls/ $0.75		
18 rolls/$1.80		

Total Cost: ____________________

(Sample Item, Connecticut State Department of Education, 2009)

Figure 3. Fifth-grade prompt.

Table 9
Fifth-Grade Student Scores on Open-ended Prompt, Spring (Post-assessment)

	Number of Students	Score				Percent Mastery
		0s	1s	2s	3s	
Pre-MLD Class (Spring 2007)	18	6 (29%)	8 (38%)	2 (10%)	5 (24%)	33%
MLD Class (Spring 2008)	23	3 (13%)	10 (43%)	6 (26%)	4 (17%)	43%

Grade 5 Results – Student Perceptions of Prompt Challenges

The students' written reflections on the prompts revealed similar trends to those reported in grade 4. An interesting, though perhaps not surprising finding, was that students who had difficulty restating the question in their own words (e.g., “I don’t know” or “It was about hamburgers and rolls”) tended to score below the mastery level; those who could restate the problem (e.g., “I think the main idea was to get the amount of food they need but don’t need to be perfect” or “To find out how many packages of each size of hamburgers and rolls that you will need to buy”) tended to perform well overall. Again, we see connections between language and mathematical performance. Being able to make sense of the written text and unpack the purpose of the problem was a challenge for many students and critical to performance.

Discussion and Implications

Promise of the Model

This small-scale implementation and study of the Mathematics Learning Discourse project offers promising results. Overall, we take these results to indicate that this model for promoting a mathematics learning discourse, grounded in the three pillars, is a productive one. Students in MLD classes demonstrated improvement in their proficiency responding to open-ended prompts beyond what was typical for their teacher or school. Analyses that focused specifically on academic language and justification also indicated improvement. Although we cannot claim that the MLD project was solely responsible for the observed differences, or that the model can be successful in all environments, urban or otherwise, at any scale, it is reasonable to assert that the teachers’ participation in the program likely had an impact on student performance and that further exploration of the model’s effectiveness is warranted. In this discussion, we reflect on some of the components of the program that seemed critical given that these were teachers working in an urban setting, as well as some of the contextual factors that seemed to support this successful case.

As we reflect on the model and its value, we are drawn to its emphasis on promoting the development of academic language, which, when coupled with justification, seems to address an important gap—especially for mathematics instruction in urban schools with linguistically diverse students. Language is a critical part of mathematics teaching, learning, and assessing. The results from the project, as well as informal conversations with the teachers, indicate that this focus was productive in bringing to the fore ideas related to language that extended *beyond* the learning of vocabulary words and low-level drilling of computation, aspects perhaps over-emphasized out of a lack of know-how for engaging students in other kinds of more cognitively and linguistically challenging mathematical activities. As noted earlier, before the project, we had the opportunity to conduct a needs assessment with approximately 10 teachers at one of the participating schools. Teachers focused their discussion of the language challenges on students not knowing enough vocabulary words. This "deficiency" was important to them because they felt that their students often missed an open-ended prompt on the state assessment because they did not know specific vocabulary like *museum* or *scuba diver*. In addition to highlighting the critical nature of this work, it indicates how significant a shift it may be for teachers of mathematics to make as they begin to see the full scope of the role of language in teaching and learning mathematics. The work of language and mathematics extends well beyond vocabulary—whether everyday or mathematical terminology—and includes engaging students in expressing mathematical ideas and core mathematical practices such as justification.

The model's focus on higher order thinking has also seemed to play a critical role. Rather than focusing narrowly on more routine skills, as is often the case in urban schools (Anyon, 1997; Keiser, 2005), and which was certainly emphasized in the schools we worked with as well, this project called attention to higher order thinking that may expand students' opportunities to learn to think mathematically. Principals and curricular directors prioritize as they try to respond to the pressures of high-stakes tests. We have heard numerous times that, in some schools, students' scores on the open-ended response items (a relatively small portion of the test in elementary grades) are so low, that teachers are told to not spend (i.e., "waste") time trying to address these areas. The "cost" to produce a measurable change is too high. The MLD project's design allowed the teachers to not only focus regularly on higher order thinking but also to have resources and a community with which to work in order to build their capacity to design and implement these lessons.

There were contextual factors that seemed to support the success of the MLD project. First, prior partnerships between the schools and university likely facilitated aspects of the project. This familiar relationship also gave us the opportunity to place three teacher education interns at these schools. The interns provided additional classroom support in the MLD classrooms, for example, organizing manipulatives, collaborating on the lesson design, or helping with responding to student work. In the context of the teachers' incredibly full work lives, these contributions helped keep the core idea of the MLD project on their radar. Second, though receiving a small stipend, these teachers volunteered for the project. Thus, they had some prior level of commitment for engaging in this kind of work. Finally, the fact that the state-mandated assessments included open-ended prompts that required language and higher order thinking provided an "in" for the project in urban schools where "adequate yearly progress" is central to the thinking of administrators; not all states include similar items on their mandated assessments.

Future Steps

Overall, this project's implementation has led us to conclude that the model for promoting a mathematics learning discourse, grounded in the three pillars, provides beginnings from which to build. Although the MLD project was a successful case, there is still much room for improvement as the majority of students still were not at mastery level. There are many challenges still to be addressed as well. For example, although not a main finding of this project, we did identify that some students struggled to articulate what they did and did not understand when reading an open-ended prompt (and subsequently deciding what mathematics to do). As part of the MLD project, we did very little with teachers regarding reading strategies and how students make sense of open-ended prompts. Rather, we focused on building background with students before engaging them in such a prompt and student production of mathematical language (verbal and written) and arguments. Such reading and comprehension activities seem important to attend to in moving forward as a key piece of students' success is interpreting from a word problem what mathematics one is to do. For example, following work done by Ratner and Epstein (2009), we could incorporate activities where a teacher asks a student to follow a think-aloud protocol as the student reads through an open-ended response, pausing to comment on what she knows, is questioning, or thinking about. This kind of activity could reveal to the teacher more about how students approach and make sense of these prompts, and how they then infer the nature of mathematical work they need to do. Such information could then guide the teacher in planning targeted instructional activities.

Another challenge is that we need to better understand the key components of teacher learning with respect to this model and the project's design so that we can scale up the intervention. In this first year, we were able to work closely with this group of teachers and we had additional school-based support with teacher education interns. Such a model is likely too human-resource intensive to replicate on a larger scale. Based on follow-up work, we suggest that it may be possible to scale up the work through a combination of professional development and grade-level team collaboration (e.g., using a modified lesson study protocol) that support teachers' sustained work on these issues with a lower level of additional personnel support (Staples & Truxaw, in press). The effort must be one of overall capacity building for a building-based community.

In conducting this work, we have become acutely aware that, as a community of education researchers, we need to develop better frameworks and methods for understanding and capturing growth in academic language in mathematics and proficiency with justification. If teachers are expected to actively promote the development of student proficiency in these areas, they need to have concrete ways to describe and understand proficiency at any moment, as well as ways to track growth over time. This issue is also one for researchers, as we need to find ways to do the same kind of documentation and description of growth.

The ALJ rubric developed for this project is perhaps a productive first step. Once adapted to a particular task, we found it useful for focusing our attention beyond students' content understandings. As we adapted the rubric to a task, we found it made us carefully examine the actual demands of the task and more fully deliberate what would be considered a complete justification and the range of ways students might approach a justification. We also had to look at the opportunities there were to express mathematical ideas and use academic language and compare that to what was required of the prompt. Although we did not directly engage teachers in using these rubrics for scoring, we have worked with teachers in PD sessions on identifying claims, warrants and evidence, and whether students articulate the links among these. These efforts seem like a productive route to continue to pursue. Future work should explore how to generate teacher- and/or student-friendly versions of such rubrics.

Thinking about change over time, however, may require different tools. We found that the set of tasks that teachers developed and implemented over the course of the year varied greatly in their demands and opportunities to demonstrate proficiency with language and justification. Consequently, it was quite challenging to understand whether variation over time was because of the task demands and opportunities or because of changes in students' proficiencies. The variation in the demands of the tasks seemed to make comparisons challenging. For example, the competitor's claim in the Graphic Design prompt can be shown false with one counterexample. This argument structure is very different from other standard justification tasks, for example, where students might be asked to generalize and offer a method for computing the perimeter of the *n*th figure in a given pattern. This latter type of task requires students to be able to express a claim about all relevant cases (a generalization) and offer a justification that likely requires coordination between a visual model and claim. These justifications make for very different mathematical work.

To our knowledge, there are no frameworks that look at mathematics academic language development and/or the development of proficiency with justification. Given the centrality of these two proficiencies to students' success in mathematics (consequential for both understanding and ability to demonstrate understanding), the development of such a framework seems quite worthwhile. To address this issue, one route to consider is the development of tasks, or sets of tasks, that would vary in their demands, but which, when taken together, paint a more complete picture than is possible with any single prompt at one point in time. These tasks potentially could allow us to better identify whether an observed change is most related to a growth in mathematical understanding, language proficiency, or proficiency with justification. These more developed measures and frameworks are also potentially valuable tools to support teachers as they work to improve students' proficiencies by providing a clearer sense of the "target" and a sense of different levels of proficiency. At this point, we seem to have an underdeveloped sense of what developing proficiency might look like over time. Thus, generating such tools might also require basic research into these areas as well.

Given the growing consensus that mastery of academic language is subject specific and critical to students' education and future success (Moje, Dillon, O'Brien, 2000; Shanahan & Shanahan, 2008), and given the influence of state assessments on teachers' classroom practice (and administrator directives), we also think it would be constructive for the state assessment system to explicitly value academic language and practices such as justification. As noted in our data, we found that the state rubrics do not explicitly attend to the students' use of academic language or expression of a justification. Yet, at the same time, the open-ended prompts on the state assessments demand these on some level: the prompts are couched in a "real-world" scenario where students must read and interpret verbal and other information in order to make sense of the situation, determine the mathematics they must do to address the question, and make some kind of claim (supported by evidence) about the correct result. There is almost an irony here. An effort has been made to not judge students explicitly along these dimensions (academic language and justification), perhaps trying to not disadvantage students unduly for language competencies, yet the prompt inherently demands great attention to language—academic (explain how you know) and otherwise (vocabulary terms)—in order for the student to begin to engage in the mathematics purposefully. Changes to the system could take the form of a rubric similar to the ALJ rubric or supporting materials that offer more explicit documentation of the demands placed on students for reading, interpreting, and responding to these open-ended prompts. It is important to stress that this call is not only about decoding or reading comprehension. Inferring from a contextual problem what mathematics one must work on to address the problem is highly mathematical work that extends beyond familiarity with each word on the page. To provide more equitable learning opportunities for students in urban schools, it is important that greater attention to the intersection between mathematics and language be given on all levels.

References

Adler, J. (1999). The dilemma of transparency: Seeing and seeing through talk in the mathematics classroom. *Journal for Research in Mathematics Education, 30*, 47–64.

Anyon, J. (1997). *Ghetto schooling: A political economy of urban educational reform*. New York: Teachers College Press.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*, 139–148.

Boaler, J., & Humphreys, C. (2005). *Connecting mathematical ideas: Middle school video cases to support teaching and learning*. Portsmouth, NH: Heinemann.

Boaler, J., & Staples, M. (2008). Creating mathematics futures through an equitable teaching approach: The case of Railside school. *Teachers College Record, 110*, 608–645.

Brenner, M. (1998). Development of mathematical communication in problem solving groups by language minority students. *Bilingual research journal, 22*(2–4), 103–128. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.5920&rep=rep1&type=pdf.

Chapin, S. H., O'Connor, C., & Anderson, N. C. (2003). *Classroom discussions using math talk to help students learn, grades 1–6*. Sausalito CA: Math Solutions Publications.

Cohen, E. G. (1994a). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1–35.

Cohen, E. G. (1994b). *Designing groupwork: Strategies for the heterogeneous classroom*. New York: Teachers College Press

Cohen, E. G., & Lotan, R. A. (1997). *Working for equity in heterogeneous classrooms: Sociological theory in practice*. New York: Teachers College Press.

Cohen, E. G., Lotan, R., & Leechor, C. (1989). Can classrooms learn? *Sociology of Education, 62,* 75–92.

Connecticut State Department of Education. (2008). *School Profiles*. Retrieved from http://www.csde.state.ct.us/public/der/ssp/index.htm.

Connecticut State Department of Education. (2009). *Mathematics Curriculum and Instruction*. Retrieved from http://www.sde.ct.gov/sde/cwp/view.asp?a=2618&q=320872.

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Buffalo, NY: Multilingual Matters.

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 2. Literacy* (2nd ed., pp. 71–83). New York: Springer.

Dalton, S., & Sison, J. (1995). *Enacting instructional conversation with Spanish speaking students in middle school mathematics*. Retrieved from http://repositories.cdlib.org/crede/ncrcdsllresearch/rr12/.

Echevarría, J., Short, D., & Powers, K. (2006). School reform and standards-based education: A model for English language learners. *Journal of Educational Research, 99*, 195–210.

Echevarría, J., Vogt, M. E., & Short, D. (2007). *Making content comprehensible for English learners: The SIOP model* (3rd ed.). Boston: Pearson/Allyn Bacon.

Echevarría, J., Vogt, M. E., & Short, D. (2010). *The SIOP model for teaching mathematics to English learners.* Boston: Pearson.

Gee, J. P. (2003). Opportunity to learn: A language-based perspective on assessment. *Assessment in Education, 10*, 27–46.

Goodlad, J. I. (1994). *Educational renewal: Better teachers, better schools*. San Francisco, Jossey-Bass.

Halliday, M. A. K. (1978). *Language as social semiotic*. London: Edward Arnold.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, *31*, 396–428.

Hiebert, J., Carpenter, T., Fennema, E., Fuson, K., Wearne, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Janzen, J. (2008). Teaching English language learners in the content areas. *Review of Educational Research, 78*, 1010–1038.

Keiser, D. L. (2005). Learners not widgets: Teacher education for social justice during transformational times. In N. Michelli & D. Keiser (Eds.), *Teacher education for democracy and social justice* (pp. 31–55). New York: Routledge.

Leonard, J., & Evans, B. R. (2008). Math links: Building learning communities in urban settings. *Journal of Urban Mathematics Education, 1*(1), 60–83. Retrieved from http://ed-osprey.gsu.edu/ojs/index.php/JUME/article/view/5/5.

Lotan, R. (2003). Group-worthy tasks. *Educational Leadership, 60*(6), 72–75.

Manouchehri, A. (2004). Implementing mathematics reform in urban schools: A study of the effect of teachers' motivation style. *Urban Education, 39*, 472–508.

Martin, D. B. (2003). Hidden assumptions and unaddressed questions in *Mathematics for All* rhetoric. *The Mathematics Educator, 13*(2), 7–11.

Michelli, N. M. (2005). Education for democracy: What can it be? In N. Michelli & D. Keiser (Eds.), *Teacher education for democracy and social justice* (pp. 3–30). New York: Routledge.

Moje, E. B., Dillon, D. R., & O'Brien, D. (2000). Reexamining roles of learner, text, and context in secondary literacy. *The Journal of Educational Research*, *93*, 165–180.

Moschkovich, J. (2002). A situated and sociocultural perspective on bilingual mathematics learners. *Mathematical Thinking and Learning, 4*(2 & 3), 189–212.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell, (Eds.), Mathematics Learning Study Committee. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Pimm, D. (1987). *Speaking mathematically: Communication in mathematics classrooms*. London: Routledge & Kegan Paul.

Ratner, A., & Epstein, S. (2009, April). *Reading math students: Math educators understanding their students as readers*. Presentation at the Inaugural CUNY Education Conference, New York, NY.

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly, 23*, 139–159.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*, 40–59.

Silver, E. A., & Stein, M. K. (1996). The QUASAR project: The "revolution of the possible" in mathematics instructional reform in urban middle schools. *Urban Education, 30*, 476–521.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher 32*(2), 3–13.

Staples, M. & Truxaw, M. (in press). Enhancing language, enhancing learning: Augmenting mathematics teachers' capacity in their linguistically diverse classrooms. *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, Columbus, OH.

Stein, M. K., Grove, B., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*, 455–488.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development*. New York: Teachers College Press.

Toulmin, S.E. (1958). *The uses of argument*. Cambridge, United Kingdom: Cambridge University Press.

Truxaw, M. P., & Staples, M. E. (2010, May). *Language-related practices to build math proficiency in linguistically diverse schools.* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Valdés, G., Bunch, G., Snow, C., Lee, C., & Matos, L. (2005). Enhancing the development of students' language(s). In L. Darling Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should know and be able to do* (pp. 126–168). San Francisco, CA: Jossey-Bass.

Vygotsky, L. S. (2002). *Thought and language* (13th ed.). Cambridge, MA: MIT Press.

Wood, T. (1999). Creating a context for argument in mathematics class. *Journal for Research in Mathematics Education, 30*, 171–191.

Wood, T., Williams, G., & McNeal, B. (2006). Children's mathematical thinking in different classroom cultures, *Journal for Research in Mathematics Education, 37*, 222–255.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*, 458–477.

Zwiers, J. (2008). *Building academic language: Essential practices for content classrooms.* San Francisco, CA: Jossey-Bass.