Gifted Child Quarterly
http://gcq.sagepub.com/

# The Impact of Advanced Curriculum on the Achievement of Mathematically Promising Elementary Students

M. Katherine Gavin, Tutita M. Casa, Jill L. Adelson, Susan R. Carroll and Linda Jensen Sheffield
*Gifted Child Quarterly* 2009 53: 188 originally published online 29 April 2009
DOI: 10.1177/0016986209334964

The online version of this article can be found at:
http://gcq.sagepub.com/content/53/3/188

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
National Association for Gifted Children

**Gifted Child Quarterly**
Volume 53 Number 3
Summer 2009 188-202

10.1177/0016986209334964
http://gcq.sagepub.com
hosted at
http://online.sagepub.com

# The Impact of Advanced Curriculum on the Achievement of Mathematically Promising Elementary Students

M. Katherine Gavin
Tutita M. Casa
Jill L. Adelson
*University of Connecticut*
Susan R. Carroll
*Words and Numbers Research, Inc*
Linda Jensen Sheffield
*Northern Kentucky University, Emerita*

***Abstract:*** The primary aim of Project M$^3$: Mentoring Mathematical Minds was to develop and field test advanced units for mathematically promising elementary students based on exemplary practices in gifted and mathematics education. This article describes the development of the units and reports on mathematics achievement results for students in Grades 3 to 5 from 11 urban and suburban schools after exposure to the curriculum. Data analyses indicate statistically significant differences favoring each of the experimental groups over the comparison group on the ITBS (Iowa Tests of Basic Skills) Concepts and Estimation Test and on Open-Response Assessments at all three grade levels. Furthermore, the effect sizes range from 0.29 to 0.59 on the ITBS Concepts and Estimation Scale and 0.69 to 0.97 on the Open-Response Assessments. These results indicate that these units, designed to address the needs of mathematically promising students, positively affected their achievement.

***Putting the Research to Use:*** To date, there is a paucity of research-based, challenging mathematics curriculum units designed specifically for mathematically promising elementary students. As a result, gifted programming for these students, if it exists within a district, often involves a collection of assorted math puzzles and problems or an above-grade-level textbook that was written for the average student. The findings from this curriculum study suggest to practitioners that mathematics curriculum units that are challenging and engaging with a focus on important math concepts and that encourage students to think and act like practicing mathematicians contribute to students' math achievement. The fact that this study was replicated with a second cohort strengthens the result. In addition, since almost 50% of the students came from economically disadvantaged backgrounds, the study illustrates that the curriculum was highly effective with this special population, while meeting the needs of all talented students.

***Keywords:*** *mathematics; curriculum; elementary; mathematically promising*

"The student most neglected, in terms of realizing full potential, is the gifted student of mathematics. Outstanding mathematical ability is a precious societal resource, sorely needed to maintain leadership in a technological world" (National Council of Teachers of Mathematics [NCTM], 1980, p. 18). Although NCTM published this quote nearly 30 years ago, progress since then has been slow. At the international level, the latest Trends in International

**Authors' Note:** The work reported here was funded in part by the Jacob K. Javits Students Education Act, grant no. S206A020006. The opinions, conclusions, and recommendations expressed in this article are those of the authors and do not necessarily reflect the position or policies of the U.S. Department of Education. The authors wish to thank all of the participating teachers and districts for their support and work with Project M$^3$. Please address correspondence to M. Katherine Gavin, Neag Center for Gifted Education and Talent Development, 2131 Hillside Road, Unit 3007, Storrs, CT 06269-3007; e-mail: kathy.gavin@uconn.edu.

Mathematics and Science Study (TIMSS, 2008) indicates that whereas more than 40% of fourth and eighth graders in Singapore and other Asian countries scored at the most advanced level, only 10% of U.S. fourth graders and 6% of eighth graders scored at this level. Results from the National Assessment of Educational Progress (NAEP, 2008) indicate that although scores continue to increase, only 6% of fourth graders and 7% of eighth graders perform at the advanced level. It is at this level that eighth graders are expected to use abstract thinking, a cornerstone of high-level mathematics. Thus, whether we look at international or national measures, our present system of mathematics education, while improving, is not serving the needs of our most capable students.

Prior to the inception of this curriculum development project, there had been a dearth of challenging, in-depth, research-based mathematics curriculum available specifically for elementary students exhibiting mathematical promise. Mathematical promise is a broadened and dynamic definition of mathematical talent or giftedness that recognizes and develops traditionally underserved students, such as those from diverse and poor backgrounds. Through a federal Jacob K. Javits research grant, Project $M^3$: Mentoring Mathematical Minds designed advanced curriculum for elementary students in an attempt to fill this void. An aim of the project was to help mathematically promising students learn more complex mathematics and achieve at internationally competitive levels.

The Project $M^3$ units were evaluated throughout the formative and summative stages of the project and results are presented here. Specifically, the purpose of this study was to determine if there were any differences in mathematics achievement between the experimental groups learning from the Project $M^3$ curriculum and a like-ability comparison group. In an effort to situate the reader, the theoretical orientation and recommendations regarding curriculum design are offered first, followed by how these were conceptualized in the Project $M^3$ units.

## Theoretical Orientation

Current mathematics education reform relies on sociocultural theory as one framework to guide its initiatives (Forman, 2003). Forman explains that sociocultural theory recognizes a connection between instruction and student learning, particularly through communication within a social context. Referencing van Oers, Forman (1996) summarized the four tenets that exemplify sociocultural theory.

> Social organizational processes are an inherent characteristic of learning—whether or not it occurs in an overtly social context. Second, learning needs to be viewed as a form of apprenticeship or a means by which novices become experts through participation in activities within a community of practice. Third, learning mathematics is a discursive activity. Fourth, learning provides the negotiation of meaning with the context of a situated activity. (pp. 116-117)

In summary, sociocultural theory frames students' learning as occurring not in isolation, but rather being influenced by the context in which the learning is taking place. Although students offer their own understandings, the teacher is in a position to mentor students with respect to practices within the discipline, in this case mathematics. In this way, the classroom becomes a *community of practice*, and communication both serves as a vehicle and sets the stage to help members of the community negotiate mathematical meaning.

## Curriculum Design Framework

Certain curriculum recommendations from both the gifted and talented and mathematics education fields can be connected to sociocultural theory. We drew from this literature, research on gifted mathematics curriculum, and recommendations from experts and major works in gifted and mathematics education to guide the development of the Project $M^3$ units.

### Contributions from the Gifted and Talented Education Field

Mathematically talented students come to know and understand mathematics differently than other students. They can use a variety of problem-solving strategies fluidly and flexibly and have a general "mathematical cast of mind" (Krutetskii, 1968/1976, p. 302). Research indicates that not only do they think differently, but their thinking actually resembles the way that professional mathematicians work (Pelletier & Shore, 2003; Sriraman, 2004). Hadamard and Polya (as cited in Sriraman, 2004), both well-respected mathematicians, believed the only difference between the work of a professional mathematician and a talented student of mathematics was in the degree of sophistication.

Encouraging students to think and act like practicing professionals is one of the hallmarks of learning promoted by experts in gifted and talented education. This philosophy is outlined in both *The Multiple Menu Model: A Practical Guide for Developing*

*Differentiated Curricula* (Renzulli, Leppien, & Hays, 2000) and in the *Curriculum of Practice*, part of the Parallel Curriculum Model (Tomlinson et al., 2002). These authors promote providing opportunities for talented students to use the skills and methodologies of the discipline they are studying. This focus on disciplinary thinking leads to a curriculum focused on solving problems that is in line with recommendations by NCTM (1989, 2000) and by leaders in the field of gifted mathematics education (Sheffield, 1994; Wheatley, 1983).

To focus on disciplinary thinking, Tomlinson et al. (2002) recommend using the Core Curriculum Model from the Parallel Curriculum. This curriculum design is built on key concepts, principles, and skills essential to the discipline. The result is a curriculum that is coherent and organized to achieve essential outcomes. The investigations in the curriculum should "cause students to grapple with ideas and questions, using both critical and creative thinking" (p. 21).

The impact of different models of mathematics curriculum for gifted students has not been fully established given the limited curriculum that is available. Tieso (2003) found that using an enhanced or differentiated curriculum with high-ability elementary students resulted in significant achievement gains compared with using a unit from the regular mathematics curriculum. Studies on different programming models of acceleration and enrichment are limited and have mixed results. For instance, Robinson, Shore, and Enersen (2007) found that acceleration enables students to cover the content efficiently. However, they caution that acceleration alone does not promote the high-level thinking that is vital and characteristic of mathematically promising students. Sowell's (1993) review of five studies focused on enrichment and found mixed results. Fourth graders outperformed the control groups on cognitive and affective measures in one study, whereas in another study fifth and sixth graders were not significantly different from the control group. Although limited studies have been conducted that focus on a combination of acceleration and achievement, significant achievement gains indicate that combination is a promising approach (Miller & Mills, 1995; Moore & Wood, 1988; Robinson & Stanley, 1989).

## Contributions From the Mathematics Education Field

NCTM has addressed the importance of considering the mathematical content students learn, how they learn it, and the environment in which this learning takes place in the design of curriculum. As Clements (2007) points out, the NCTM *Standards* "were created by a dialectical process among many legitimate stakeholders and thus serve as a valuable starting point" (p. 40) in helping establish the educational goals of the mathematics curriculum. They identified five major content areas to be studied across all grades: algebra, data analysis and probability, geometry, measurement, and number and operations. Recently, NCTM addressed the critique of the U.S. mathematics curriculum as being "a mile wide and an inch deep" (Fuson, 2004; Schmidt, Wang, & McNight, 2005) with the publication of the *Curriculum Focal Points* (2006). This document stresses the depth of student learning over the coverage of numerous content areas.

The curriculum recommendations included in the aforementioned publications built on a foundation established by NCTM's *Curriculum and Evaluation Standards for School Mathematics* (1989). In this seminal publication, NCTM outlined some guiding curriculum principles. First, the level and depth at which students come to understand mathematical concepts is more highly regarded than the number of skills they obtain. Second, affective considerations need to be considered in curriculum development. That is, the curriculum should "build beliefs about what mathematics is, about what it means to know and do mathematics, and about children's view of themselves as mathematics learners" (pp. 16-17). These principles should be manifested in curriculum that has a conceptual orientation, encourages students to be actively engaged with mathematics, emphasizes students' developing reasoning abilities, and addresses content beyond arithmetic, among others (NCTM, 1989).

The 2000 NCTM *Standards* address not only *what* students should learn through the content standards, but also *how* they learn it via the process standards. Recently, Boix Mansilla, and Gardner (2008) have continued to promote learning the discipline and disciplinary thinking rather than simply subject matter: "The goal of this approach is to instill in the young the disposition to interpret the world in the distinctive ways that characterizes the thinking of experienced disciplinarians" (pp. 14-15). In line with gifted and talented recommendations (Renzulli et al., 2000; Tomlinson et al., 2002), the process standards present teachers with strategies to engage their students with mathematics in ways that are similar to those of practicing mathematicians engaged with the discipline (Sriraman, 2004). These processes include communication, connections, reasoning, representation, and

**Table 1**
**A Priori Foundations Stages in the Development of the Project M$^3$ Units**

| A Priori Foundation | Project M$^3$ Development |
|---|---|
| 1. Subject matter | The curriculum authors, all national or state-level leaders in gifted education and mathematics education, primarily relied on NCTM content standards as a guide. They referred to Connecticut, Kentucky, and Massachusetts state standards (the authors' home states) and NSF-developed curricula to identify more specific objectives. Mathematicians, mathematics educators, gifted educators, mathematics specialists, teachers, and professional development leaders served as content reviewers prior to field testing. |
| 2. Pedagogy | The authors reviewed gifted education literature to identify and recommend best curriculum practices. Sociocultural theory served to merge mathematics education and gifted education recommendations. |
| | Enrichment teaching and learning strategies from the gifted education field, in particular differentiation and a focus on student as practicing professional, were used to support and develop content knowledge. The authors also embedded the NCTM process standards particularly communicating to support students' problem solving and reasoning within the curriculum. |

Note: NCTM = National Council of Teachers of Mathematics, NSF = National Science Foundation.

problem solving and are inextricably linked. For instance, students use reasoning as they solve problems and then communicate their reasoning using a variety of representations.

The mathematics education field gives considerable attention to the context in which students learn, particularly in an effort to support the execution of the process standards. Leaders have noted that the classroom environment should emphasize how students understand and come to know the mathematics (Wood, 1999), support student access to discussions through established norms (Hiebert et al. 1997), and encourage and accommodate the exchange of multiple perspectives by treating misconceptions as opportunities for learning (Hiebert et al., 1997; Kazemi, 1998; Mewborn & Huberty, 1999). Students should take on roles, particularly during discussions, by listening and responding to others through speculation, investigation of conjectures, and presentation of viable solutions. They also should convince themselves and others of the validity of their ideas and depend on evidence grounded in mathematics to determine the validity of ideas (NCTM, 1991). This is in line with the tenets of sociocultural theory that provided a framework for the curriculum. The development of the Project M$^3$ units addressed the concepts addressed in the mathematics education literature—content, process, and learning environment—and the specific components are detailed next.

## Development of the Project M$^3$ Units

A partnership of gifted and talented educators, mathematicians, and mathematics educators collaborated to write the 12 Project M$^3$ units and meld the recommendations set forth in the gifted and talented and mathematics education fields. The development of the Project M$^3$ units paralleled the a priori foundations phases from the Curriculum Research Framework more recently proposed by Clements (2007). *A prior foundations* entail phases when "extant research is reviewed and implications for the nascent curriculum development effort [is] drawn" (p. 42). Table 1 summarizes the actions taken in the development of the Project M$^3$ curriculum units as it maps onto Clements's a priori foundations. A more thorough description of the particular features of the units follows.

In an effort to allow for flexibility in implementation, the Project M$^3$ units were designed as individual units rather than a complete curriculum. There are a total of 12 units, with 4 units at each of 3 levels primarily designed for students in Grades 3, 4, and 5. Each unit addresses important mathematical ideas from one of the NCTM content strands, which were grouped accordingly: (a) algebra, (b) data analysis or probability, (c) geometry or measurement, and (d) number and operations. The content is accelerated one to two grade levels, and students investigate the mathematics in-depth. The process standards are embedded throughout the units in an effort to position students as practicing mathematicians. Although the units emphasize verbal and written communication centered on important mathematical ideas, students also regularly use the NCTM processes of problem solving, reasoning, making connections, and creating and using representations. The tasks demand high-level thinking and the creation of products that encourage students to extend what they have learned in various ways, such as games and culminating projects.

In line with a sociocultural perspective (Forman, 1996; 2003), the Project M$^3$ units also provide

**Table 2**
**Connections Between the Project M$^3$ Units and Literature Recommendations**

| Unit Features | Connection to the Literature |
|---|---|
| Important mathematical ideas: Students think in-depth about the essential concepts for a particular content area | Conceptual orientation (NCTM, 1989); Content standards (NCTM, 2000); In-depth investigations (NCTM, 1989, 2006); Accelerated and enriched content (e.g., Robinson & Stanley, 1989; Moore & Wood, 1988; Miller & Mills, 1995); Core curriculum (Tomlinson, et al. 2002) |
| Differentiation: Different levels of support and challenge are provided, including (a) *Hint Cards* for students needing support; (b) *Think Deeply Questions* for most students; and (c) *Think Beyond Cards* for students needing further challenge | Differentiated curriculum for different levels of talent (Tieso, 2003; Tomlinson, 1995); Written communication as a mathematical process (NCTM, 2000) |
| Projects and/or culminating activities: Student projects address the big ideas and focus on students as practicing mathematicians | Thinking like a practicing mathematician (Boix Mansilla & Gardner, 2008; Chazan & Ball, 1995; Renzulli, Leppien, & Hays, 2000; Sriraman, 2004; Tomlinson et al., 2002); In-depth investigations (NCTM, 1989; 2006); Process standards (NCTM, 2000); Active engagement with mathematics (NCTM, 1989) |
| Verbal discourse: Talk moves (Chapin, O'Connor, & Anderson, 2003), particularly agree/disagree and why, establish a community of practice to make meaning of mathematics | Verbal communication as a mathematical process (NCTM, 2000); Thinking like a practicing mathematician (Boix Mansilla & Gardner, 2008; Chazan & Ball, 1995; Renzulli, Leppien, & Hays, 2000; Sriraman, 2004; Tomlinson et al., 2002); Developing reasoning abilities (NCTM, 1989) |
| Classroom environment: Classroom Rights and Obligations guide the social norms | Classroom environment (Hiebert et al., 1997; Kazemi, 1998; Mewborn & Huberty, 1999; Wood, 1999) |

Note: NCTM = National Council of Teachers of Mathematics.

strategies for teachers to address the culture of the classroom in an effort to engage students as practicing mathematicians who commonly reason and justify their ideas. Teachers reinforce the "Classroom Rights and Obligations" that outline students' expected behaviors. Students have the right to ask questions, to make a contribution to an attentive and responsive audience, to be treated respectfully, and to have their ideas discussed. Students also are obligated to speak loudly enough for others to hear, to listen to others in order to understand, and to agree or disagree with the speaker's comments and explain why. The Rights and Obligations serve to support and are supported by the classroom discourse. Specifically, teachers are shown how to implement Chapin, O'Connor, and Anderson's (2003) talk moves to facilitate discussions. They learn how to *revoice* student contributions, have students *repeat/rephrase* one another's ideas, encourage them to *agree/disagree* and explain *why* using mathematically valid evidence, have students *add on* additional perspectives, and use *wait time* to encourage more contributions. In particular, the *agree/disagree* and *why* talk move embodies the design of the questions students respond to in writing. The Think Deeply questions are intended to be the heart and soul of each lesson as they address a concept tied directly to an important mathematical idea. Students frequently have to justify their mathematical position by explaining their reasoning using evidence. In Table 2, we provide a brief overview of how the curriculum unit features are connected to the literature recommendations in gifted and mathematics education.

## Method

Although recommendations from the gifted and mathematics education fields are impingent to the development of new curriculum, implementing these into the curriculum design is not sufficient to determine their efficacy; the curriculum needs to be evaluated to ensure that gains in student achievement are imminent (VanTassel-Baska, Zuo, Avery, & Little, 2002). The first level of evaluation of the units involved a content analysis by gifted and mathematics education experts and teachers. Written feedback was gathered, analyzed, and used in the revision of the units used for the field test. Then, we examined the effectiveness of the Project M$^3$ units using a quasi-experimental design focused on students' mathematics achievement.

## Selection of Schools and Teachers

The researchers and curriculum authors intended the Project $M^3$ units to challenge students from all backgrounds, including those from lower and higher socioeconomic districts. The major emphasis of the U.S. Department of Education Javits program is on serving students traditionally underrepresented in gifted and talented programs, particularly economically disadvantaged, limited English proficient, and students with disabilities, to help reduce the serious gap in achievement among certain groups of students at the highest levels of achievement. Higher socioeconomic districts were included to ensure that this would not be considered a compensatory curriculum. This is in agreement with Clements (2007), who makes note of "the importance of representative populations when the structure and content of curricula are being formed" (p. 47).

Schools from urban and suburban areas of Connecticut and Kentucky agreed to participate in the study for 4 years, with teachers in each grade level (3-5) committing to participate for 2 consecutive years within this time frame. In most schools, students left their homerooms during their regularly scheduled mathematics period to participate in the intervention. Teachers were not randomly assigned as they also had to commit to a 2-week professional development institute in the summer prior to the first implementation at their grade level. The first 10 schools that agreed to these conditions participated in the study. A total of 11 schools participated (two elementary schools fed into one middle school in Grade 5). Nine were in Connecticut and two in Kentucky; with seven in urban settings, and four in suburban districts.

## Sample of Students

*Identification.* In an effort to target underrepresented groups and support a diverse sample, a more inclusive group of students, using a broadened definition of mathematical talent, was identified to participate in the study. NCTM defines the group of students with high ability in mathematics as *mathematically promising*. The NCTM Task Force on Mathematically Promising Students identifies mathematical promise as "a function of ability, motivation, belief, and experience or opportunity." They also state that students who possess this have a "large range of abilities and a continuum of needs that should be met" (Sheffield, 1999, p. 310). The researchers used this broadened definition of mathematically talented students in selection of the research sample and in the development of the Project $M^3$ units.

The researchers followed a strict identification procedure based on exemplary practices in gifted education. To ensure comparability of groups, they identified the experimental and comparison groups in exactly the same way in the same schools. The National Research Council (Confrey, 2006) strongly recommends this methodology for use in curriculum research studies. The identification procedure included the use of multiple measures to identify mathematical promise in students (Gavin & Adelson, 2007; Sheffield, Bennett, Berriozabal, DeArmond, & Wertheimer, 1999; Sowell, Bergwell, Zeigler, & Cartwright, 1990). The instruments used to identify students were the Naglieri Nonverbal Ability Test (NNAT; Harcourt Brace Educational Measurement, 1997; KR-20 reliability = .83), the Mathematics Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS Math Scale; Cronbach $\alpha$ = .98; Gavin, 2005), classroom performance, other standardized tests given by the district, and other pertinent information teachers shared about students. The NNAT includes items on pattern completion, reasoning by analogy, serial reasoning, and spatial visualization. Because these processes were some of the components of the Project $M^3$ curriculum, the researchers felt strong performance on this test would help identify appropriate participants. In addition, this assessment is appropriate for identifying students from diverse cultural and language backgrounds and those with learning differences because it uses pictures without any words (Harcourt Brace Educational Measurement, 1997). At the same time, all the grade-level teachers completed the SRBCSS Math Scale on the top half of their class. This scale rated students on how well they exhibited characteristics of mathematically talented students such as using creative and unusual ways to solve math problems, displaying a strong number sense, and frequently solving math problems abstractly, without the need for concrete materials. Teachers also filled out a recommendation form that included classroom performance, standardized test scores, and any other relevant information they wished to share about the students.

In aiming for an approximate class size of 20 students per school, researchers used local norms to identify students. If there was inconsistency between measures, researchers contacted teachers to discuss the individual student. Often this inconsistency was the result of selecting a student who had not been identified by their teacher as being in the top half of their class yet performed at high levels (~85th percentile or above) on the NNAT. More often than not,

**Table 3**
**Student Demographics for Experimental and Comparison Groups**

| Group | *n* | Gender (%) | Ethnicity/Race (%) | Eligible for Meal Subsidy (%) |
|---|---|---|---|---|
| Experimental Group I | 193 | Males (53)<br>Females (47) | Caucasian (54)<br>Multiethnic/racial (46) | 46 |
| Experimental Group II | 177 | Males (53)<br>Females (47) | Caucasian (53)<br>Multiethnic/racial (47) | 46 |
| Comparison Group | 211 | Males (55)<br>Females (45) | Caucasian (51)<br>Multiethnic/racial (49) | 47 |
| School profiles 2002-2003[a] | | | Caucasian (52)<br>Multiethnic/racial (48) | 52 |
| School profiles 2003-2004[b] | | | Caucasian (51)<br>Multiethnic/racial (49) | 51 |

a. 2002-2003 is the school year in which Experimental Group I and Comparison Group were identified. Gender information not available from school profile.
b. 2003-2004 is the school year in which Experimental Group II was identified. Gender information not available from school profile.

**Table 4**
**Comparison of Each Experimental Group With the Comparison Group Prior to Intervention**

| Measure | Group | *N* | Mean (*SD*) | *t* (df) | *p* |
|---|---|---|---|---|---|
| Naglieri Nonverbal Ability Test | Comparison | 211 | 114.48 (12.65) | | |
| | Experimental I | 187 | 116.69 (14.36) | 1.63 (396) | .10 |
| | Experimental II | 182 | 115.91 (13.96) | 1.07 (391) | .29 |
| SRBCSS Math Scale | Comparison | 181 | 27.14 (5.94) | | |
| | Experimental I | 168 | 26.39 (5.61) | 1.21 (347) | .23 |
| | Experimental II | 164 | 49.61 (6.71) | 32.98 (343) | < .0001 |

Note: SRBCSS = Scales for Rating the Behavioral Characteristics of Superior Students.

teachers took a second look at the student and agreed that this student might have hidden potential in terms of reasoning ability. (See Gavin, 2005, to learn more about the identification measures.)

*Experimental and comparison groups.* There were two experimental groups. Students in Experimental Group I were selected during the first year of program implementation (2003), and students in Experimental Group II were selected during the following year (2004). The Comparison Group was a sample from the same schools identified in the same way as the experimental groups and was selected and assessed the year prior to the intervention in the schools (2002). They did not receive the intervention. This helped assure no diffusion of treatment from the experimental groups to the comparison group.

In addition to the identical procedures used to identify students for the experimental and comparison groups, the demographic characteristics reported in Table 3 confirm the similar profiles of the three groups and their comparability to the overall school populations from which the cohorts were chosen. Unlike studies that compromise external validity with homogeneous samples from largely suburban schools, Project M³ had diversified subjects in each of the three groups. The profiles on gender, ethnicity/race, and household income as indicated by "eligible for a meal subsidy" reflect a broad population of students. In addition, these demographic statistics are comparable with those of the entire school population from which the sample was chosen. Experimental Group I and the Comparison Group were chosen from the same schools and identified in the same year (2002-2003). During this year, the school profiles for ethnicity/race were 52% Caucasian and 48% multiethnic/racial. A total of 52% of the students were eligible for a meal subsidy. In 2003-2004, Experimental Group II was identified. The demographic statistics for that year for the school population were again similar to the sample chosen. The school profiles showed 51% of the population was Caucasian, 49% was multiethnic/racial, and 51% of the students were eligible for a meal subsidy.

As shown in Table 4, students in the comparison group had comparable scores to both experimental groups on the NNAT. Whereas students in both the comparison group and Experimental Group I received comparable ratings from their teachers on the SRBCSS Math Scale, students in Experimental Group II received statistically significantly higher ratings than students in the comparison group. Through observation and discussion with teachers, the researchers learned that teachers became more aware of characteristics of mathematically promising students and of problems, opportunities, and questions that allow students to demonstrate their talent potential as a result of the implementation of the project the previous year. In fact, many of the third-grade teachers commented to the researchers that after the first year of the project they further defined the characteristics of mathematically promising students for the second-grade teachers and discussed which students would benefit from the program. Thus, taking this caveat into consideration together with the similar scores on the NNAT, it is reasonable to assume that both experimental groups and the comparison group were comprised of like-ability students.

## Intervention

Prior to the implementation of the Project $M^3$ units, teachers attended a 2-week professional development summer institute during which they learned about the philosophy, teaching strategies, and content of the units. During the school year, they received 1 day of training prior to the implementation of each unit. Each unit spanned approximately 6 weeks of instructional time, and teachers implemented three or four of the grade-level units for approximately one-half of each school year. During the remainder of the school year, teachers compacted the regular curriculum and taught objectives not addressed in the Project $M^3$ units. In the second year of the project (2003), Experimental Group I began studying the units in Grade 3 and continued in Grades 4 and 5. In the third year of the project (2004), Experimental Group II began studying the units in Grade 3 and also continued through Grade 5.

During the implementation phase, Project $M^3$ team members visited each of the experimental classrooms once a week. The purpose of these visits was twofold. First, these visits provided fidelity of implementation checks in each of the classrooms. The Project $M^3$ team could assess whether or not the material in the unit was being taught and whether or not it was being taught in the way it was intended to be. In addition, teachers were required to keep a written record of the number of days each lesson was taught and how they used the different unit components. While in the classroom, the project team also could formatively assess the impact of individual lessons on student participation and understanding. Second, the visits served as additional time to work with teachers on lesson planning and gain their feedback on a weekly basis about the curriculum and instructional strategies in the program. If teachers were not following the prescribed teaching strategies or sequence, the team member could model the intended approach in the classroom and also find out where the difficulties were. This helped the authors revise the content, pedagogy, format, and mathematics background in the teacher guide.

## Research Design

We examined whether there was a difference in mathematics achievement between mathematically promising students exposed to the intervention and a comparison group of students of similar abilities and backgrounds. Because the sample was a restricted one (limited only to mathematically promising students) and drawn from urban as well as suburban schools, randomized control trials were not practical. The potential for attrition, as well as possible scheduling changes over the 3-year period of the intervention, presented a threat to internal validity. However, methodological rigor was built-in to the research design to ensure that the results were authentic. To do this, the research hypothesis was tested twice, an internal replication. This design decision addressed the issues raised in evaluating curricular effectiveness (Clements, 2007; Collins et al., 2004; Confrey, 2006; Kelly, 2004). It also addressed the concerns of the What Works Clearinghouse and the National Research Council that no single study should be used to make policy decisions (Confrey, 2006).

## Data Collection and Analysis

As Confrey (2006) recommends, multiple measures of mathematics achievement were used to evaluate the effectiveness of a curricular intervention. The Concepts and Estimation Test of the Iowa Tests of Basic Skills (ITBS), a norm-referenced standardized assessment, was selected to measure the difference between the experimental and comparison groups at the end of the third, fourth, and fifth grades. We chose this scale because it reflected the content addressed in Project $M^3$ because it "focus[es] on numeration, properties of number systems, and number

**Table 5**
**Descriptive Statistics for Experimental and Control Groups**

| Grade and Variable | Experimental Group I | | | Experimental Group II | | | Comparison Group | | |
|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *N* | *M* | *SD* | *N* | *M* | *SD* | *N* |
| Third | | | | | | | | | |
| ITBS | 200.63 | 23.88 | 185 | 203.52 | 16.45 | 172 | 194.42 | 20.35 | 211 |
| Open-response | 8.74 | 2.36 | 184 | 8.40 | 2.30 | 172 | 6.33 | 2.38 | 208 |
| Fourth | | | | | | | | | |
| ITBS | 226.24 | 20.70 | 178 | 224.66 | 19.69 | 156 | 214.06 | 20.95 | 180 |
| Open-response | 10.11 | 3.51 | 177 | 9.91 | 3.23 | 159 | 6.49 | 3.22 | 180 |
| Fifth | | | | | | | | | |
| ITBS | 241.62 | 22.18 | 163 | 246.42 | 21.50 | 142 | 233.18 | 22.96 | 147 |
| Open-response | 7.64 | 2.66 | 162 | 8.25 | 2.33 | 143 | 5.73 | 2.50 | 147 |

Note: ITBS = Iowa Tests of Basic Skills.

sequences; fundamental algebraic concepts; and basic measurement and geometric concepts . . . and probability and statistics" (Hoover et al., 2003, p. 38). As such, it met the test of "curricular validity" identified by the National Research Council as critical (Confrey, 2006, p. 203). As corroborating assessments, there were open-response questions addressing major unit concepts and derived from released items on the NAEP and TIMSS assessments. Students completed the open-response questions at the end of each grade. To eliminate the diffusion of the treatment to the comparison group, the comparison group students were tested *prior* to the grade-level Project $M^3$ intervention so that rival hypotheses would be reduced.

## Results

To investigate differences in mathematics achievement as measured on a traditional assessment (the ITBS Concepts and Estimation Test) and on an Open-Response Assessment across experimental and comparison groups, we conducted a series of 2-level multilevel models using hierarchical linear modeling version 6.06 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004). The dependent variables of interest were scores on the Concepts and Estimation section of the ITBS and scores on the Open-Response Assessment for each grade level. For the Open-Response Assessment, we combined the total scores on all items that were administered to all three groups for each grade level without weighting any items. Table 5 contains descriptive statistics for the outcome measures used in this study.

Although data were collected at the student level, we were interested in testing classroom-level effects. Level 1 contained mathematics outcome scores for students; Level 2 contained classroom information, that is, particular experimental or control group and school. The independent variable of greatest interest, exposure to Project $M^3$ curriculum, included three conditions—the Project $M^3$ Experimental Group I, the Project $M^3$ Experimental Group II, and the Comparison Group. Because we had three groups, we used two dummy codes—M3_ExpI (Experimental Group I was coded 1; the other groups were coded 0) and M3_ExpII (Experimental Group II was coded 1; the other groups were coded 0), and we entered these two variables at Level 2. For the other Level 2 variable, school, we created nine dummy codes for the 10 cohorts of students. Given the small Level 2 sample size, we used restricted maximum likelihood estimation (Raudenbush & Bryk, 2002).

Normality of Level 1 residuals and the homogeneity of Level 1 variances are standard assumptions of hierarchical linear modeling version 6.06 . We set the alpha level for the test of homogeneity of variance to .02 because this assumption is powerful and extremely sensitive to nonnormality (Raudenbush & Bryk, 2002). Among the basic residual analyses we conducted were examination of the normality of the Level 1 residuals. All of the outcome scores exhibited slight departures from normality (with some skew or kurtosis values greater than $|0.25|$, although none greater than $|1.00|$), as did the Level 1 residuals for the ITBS concepts and estimation and the Open-Response Assessment in Grade 3. For all three grade levels, the ITBS concepts and estimation outcome scores exhibited heterogeneous Level 1 variances ($p < .02$), but the

Open-Response Assessment outcome scores exhibited homogeneous Level 1 variances (Grade 3, $\chi^2(29) = 26.59$, $p = .157$; Grade 4, $\chi^2(29) = 46.52$, $p = .021$; Grade 5, $\chi^2(29) = 30.63$, $p = .383$). The departures from normality for outcomes and some residuals seemed a plausible reason that, in some of the cases, we rejected the assumption of homogeneity of variances. According to Raudenbush and Bryk (2002), if we were to use full information maximum likelihood estimation techniques, which are required to model the heterogeneity of variances explicitly, we would have biased estimates of the variance components because we had a relatively small number of classes (Level 2 units). Therefore, we did not model the heterogeneity of variances in any outcome scores explicitly. Instead, we used the robust standard errors, which are considered more robust to violations of normality and homogeneity than are the conventional errors.

We began analyzing each of the six outcomes by estimating a baseline model with no predictors at either level so that we could estimate the intraclass correlation (ICC), a measure of the proportion of variance at the school level in relation to the total variance. For all three grade levels of the ITBS concepts and estimation test, the ICC was about .30 (Grade 3 = .309, Grade 4 = .318, Grade 5 = .293). This indicates that about 30% of the variance in ITBS concepts and estimation scores at each grade level lay between classes. The results of each baseline model for the ITBS concepts and estimation scores are in Table 6. The ICCs for the Open-Response Assessment at each grade level were somewhat more variable. For third grade, the ICC was .352; for fourth grade, it was .502; and for fifth grade, it was .364. This indicates that between 35% and 50% of the variance in Open-Response Assessment scores at each grade level lay between classes. Table 7 has the results of each Open-Response Assessment baseline model.

Given that we were interested in the effects of the Project $M^3$ intervention and did not include any Level 1 covariates, we next estimated the full Level 2 models, which included school cohort (nine dummy codes) and Project $M^3$ group (two dummy codes) at Level 2. Table 6 displays the results of the three full models for the ITBS concepts and estimation, and Table 7 displays the results of the three full models for the Open-Response Assessment. Because of the coding system, the intercept ($\gamma_{00}$) represented the predicted outcome score for a student in that grade at School 1 (all school dummy codes = 0) in the comparison group (M3_ExpI = 0 and M3_ExpII = 0). The coefficient for the nine school cohorts ($\gamma_{03}$ to $\gamma_{011}$) represented the differential between comparison group scores for students at the other nine schools and School 1. The coefficient for M3_ExpI ($\gamma_{01}$) represented the differential for a student who participated in Project $M^3$ Experimental Group I at the same school, and the coefficient for M3_ExpII ($\gamma_{02}$) represented the differential for a student who participated in Project $M^3$ Experimental Group II at the same school. This information can be used to determine the predicted score for different students. For instance, the predicted ITBS concepts and estimation score for a student in Grade 3 ($\gamma_{00} = 181.17$) at School 3 ($\gamma_{04} = 24.66$) who was in Experimental Group II was 213.79 (181.17 + 24.66 + 7.96), whereas the predicted score for a third grader in the same school but in the Comparison Group was 205.83.

Of greatest interest to us were the main effects of M3_ExpI and of M3_ExpII. For all three grade levels, both Experimental Group I and Experimental Group II had statistically significantly ($p < .01$) higher scores on the ITBS concepts and estimation. As shown in Table 8, the Cohen *d* effect sizes on this test ranged from 0.29 to 0.59, which are small to medium effect sizes (Cohen, 1992). The results from the ITBS concepts and estimation test, a standardized, multiple-choice assessment, were corroborated by those obtained on the Open-Response Assessment that consisted of items from released TIMSS and NAEP assessments. Both Experimental Group I and Experimental Group II scored statistically significantly higher ($p < .001$) on the Open-Response Assessment at all three grade levels. The Cohen's *d* effect sizes on this assessment, which also are displayed in Table 8, ranged from 0.69 to 0.97, which are medium to large effect sizes (Cohen, 1992). These results indicate that both of the Project $M^3$ experimental groups, on average, outperformed comparison students on both the ITBS concepts and estimation and the Open-Response Assessment in Grades 3, 4, and 5.

## Discussion and Implications

The main purpose of this research was to measure the efficacy of curriculum units that were designed for mathematically promising students and based on comprehensive principles from the fields of mathematics education and gifted and talented education. Results from the analyses of data show that in all three grades students in the experimental groups consistently had statistically significant gains over similarly identified students in the comparison group on standardized achievement tests (the ITBS Concepts and Estimation Test) and on open-response items from the TIMSS and NAEP assessments. These findings

**Table 6**
**Summary of REML Parameter Estimates for Two-Level Model of ITBS Concepts and Estimation, Grades 3, 4, and 5**

| | Third Grade | | | | Fourth Grade | | | | Fifth Grade | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Unconditional Model | | Full Model | | Unconditional Model | | Full Model | | Unconditional Model | | Full Model | |
| Parameter | Parameter Estimate | *SE* | Parameter Estimate | *SE* | Parameter Estimate | *SE* | Parameter Estimate | *SE* | Parameter Estimate | *SE* | Parameter Estimate | *SE* |
| Fixed effect | | | | | | | | | | | | |
| Intercept ($\gamma_{00}$) | 200.28*** | 1.99 | 181.17*** | 4.66 | 220.74*** | 2.25 | 201.72*** | 2.24 | 239.33*** | 2.46 | 219.64*** | 2.98 |
| M3_ExpI ($\gamma_{01}$) | | | 5.33** | 1.78 | | | 12.04*** | 1.52 | | | 7.42** | 2.00 |
| M3_ExpII ($\gamma_{02}$) | | | 7.96** | 2.05 | | | 9.13*** | 1.66 | | | 13.17*** | 2.45 |
| School 2 ($\gamma_{03}$) | | | 15.83** | 5.22 | | | 14.76*** | 2.59 | | | 18.97*** | 3.40 |
| School 3 ($\gamma_{04}$) | | | 24.66*** | 4.91 | | | 29.04*** | 3.21 | | | 29.10*** | 4.35 |
| School 4 ($\gamma_{05}$) | | | –1.68 | 5.22 | | | –4.00 | 4.55 | | | –3.53 | 4.81 |
| School 5 ($\gamma_{06}$) | | | 9.58 | 4.58 | | | 2.82 | 2.30 | | | 9.24** | 2.78 |
| School 6 ($\gamma_{07}$) | | | 11.73* | 4.59 | | | 6.33* | 2.46 | | | 7.04* | 2.65 |
| School 7 ($\gamma_{08}$) | | | 21.79*** | 4.85 | | | 17.92*** | 2.45 | | | 20.27*** | 2.82 |
| School 8 ($\gamma_{09}$) | | | 27.05*** | 4.46 | | | 27.70*** | 3.37 | | | 29.12*** | 2.72 |
| School 9 ($\gamma_{010}$) | | | 23.77*** | 5.18 | | | 19.02*** | 3.32 | | | 15.21* | 5.20 |
| School 10 ($\gamma_{011}$) | | | 13.16* | 5.64 | | | 3.20 | 2.61 | | | –0.06 | 5.82 |
| Variance estimate | | | | | | | | | | | | |
| Level-1 variance ($\sigma^2$) | 230.16 | | 230.16 | | 285.26 | | 285.03 | | 364.23 | | 364.03 | |
| Intercept variance ($\tau_{00}$) | 102.82*** | | 14.64* | | 133.22*** | | 1.58 | | 150.70*** | | 12.19 | |
| Deviance (number of REML parameters) | 3294.73 (2) | | 3200.49 (2) | | 3379.74 (2) | | 3272.43 (2) | | 3472.36 (2) | | 3370.53 (2) | |

Note: REML = restricted likelihood estimation; ITBS = Iowa Tests of Basic Skills. M3_ExpI and M3_ExpII (Project $M^3$ Experimental Groups I and II) are indicators of treatment group and are dummy coded 0 for comparison and 1 for participation in the respective Experimental Group. Nine dummy codes were created to represent the 10 school cohorts.
*$p < .05$. **$p < .01$. ***$p < .001$.

**Table 7**
**Summary of REML Parameter Estimates for Two-Level Model of Open-Response Assessment, Grades 3, 4, and 5**

| | Third Grade | | | | Fourth Grade | | | | Fifth Grade | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Unconditional Model | | Full Model | | Unconditional Model | | Full Model | | Unconditional Model | | Full Model | |
| Parameter | Parameter Estimate | *SE* | Parameter Estimate | *SE* | Parameter Estimate | *SE* | Parameter Estimate | *SE* | Parameter Estimate | *SE* | Parameter Estimate | *SE* |
| Fixed effect | | | | | | | | | | | | |
| Intercept ($\gamma_{00}$) | 7.78*** | 0.30 | 5.59*** | 0.19 | 8.75*** | 0.51 | 4.97*** | 0.39 | 7.02*** | 0.33 | 4.76*** | 0.63 |
| M3_ExpI ($\gamma_{01}$) | | | 2.52*** | 0.29 | | | 3.70*** | 0.37 | | | 1.92*** | 0.28 |
| M3_ExpII ($\gamma_{02}$) | | | 2.24*** | 0.27 | | | 3.54*** | 0.38 | | | 2.48*** | 0.34 |
| School2 ($\gamma_{03}$) | | | 0.74** | 0.18 | | | 3.61** | 0.96 | | | 2.08* | 0.71 |
| School3 ($\gamma_{04}$) | | | 2.21*** | 0.25 | | | 3.40*** | 0.43 | | | 2.09** | 0.58 |
| School4 ($\gamma_{05}$) | | | –0.98** | 0.33 | | | –1.64 | 0.94 | | | –1.67* | 0.76 |
| School5 ($\gamma_{06}$) | | | 0.55** | 0.16 | | | –0.56 | 0.40 | | | 0.52 | 0.60 |
| School6 ($\gamma_{07}$) | | | –0.14 | 0.30 | | | 0.03 | 0.39 | | | 0.14 | 0.72 |
| School7 ($\gamma_{08}$) | | | 1.12 | 0.60 | | | 2.61*** | 0.51 | | | 1.15 | 0.61 |
| School8 ($\gamma_{09}$) | | | 1.38*** | 0.16 | | | 4.13*** | 0.40 | | | 1.44* | 0.61 |
| School9 ($\gamma_{010}$) | | | 1.90** | 0.51 | | | 2.88*** | 0.35 | | | 2.51** | 0.64 |
| School10 ($\gamma_{011}$) | | | –0.88* | 0.30 | | | –1.15 | 0.69 | | | –0.56 | 0.84 |
| Variance estimate | | | | | | | | | | | | |
| Level-1 variance ($\sigma^2$) | 4.38 | | 4.35 | | 7.24 | | 7.24 | | 4.99 | | 4.99 | |
| Intercept variance ($\tau_{00}$) | 2.38*** | | 0.26* | | 7.31*** | | 0.54** | | 2.86*** | | 0.33* | |
| Deviance (number of REML parameters) | 1750.30 (2) | | 1692.08 (2) | | 1963.26 (2) | | 1887.14 (2) | | 1803.15 (2) | | 1745.32 (2) | |

Note: REML = restricted likelihood estimation. M3_ExpI and M3_ExpII (Project $M^3$ Experimental Groups I and II) are indicators of treatment group and are dummy coded 0 for comparison and 1 for participation in the respective Experimental Group. Nine dummy codes were created to represent the 10 school cohorts.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

**Table 8**
**Cohen's *d* Effect Sizes for All Outcome Measures**

| | ITBS Concepts and Estimation | | | Open-Response Assessment | | |
|---|---|---|---|---|---|---|
| Experimental group | Grade 3 | Grade 4 | Grade 5 | Grade 3 | Grade 4 | Grade 5 |
| I | 0.29 | 0.59 | 0.33 | 0.97 | 0.97 | 0.69 |
| II | 0.44 | 0.45 | 0.58 | 0.86 | 0.93 | 0.89 |

Note: ITBS = Iowa Tests of Basic Skills.

are consistent with results of other studies conducted on reform-based curricula (see Clements, 2007; Senk & Thompson, 2003). Features of the study design that strengthen the results include a comparison group of like-ability students with similar demographics from the same schools as the experimental groups and who were tested prior to any grade-level intervention, use of multiple and varied assessments to measure math achievement, and replication of the implementation at each of the three grade levels with a second experimental group. These results provide initial "proof of concept" (NCTM, 2007, p. 2) support for efficacy of the Project M$^3$ curriculum units on the achievement of mathematically promising students. In doing so, they provide an estimate of effectiveness with on-site supervision during the implementation of the curriculum that insured fidelity of treatment.

The small number of classrooms, the weekly presence of project staff, and other professional development offerings limit generalizability. Nevertheless, the positive results of this study suggest directions for future research. Disaggregation by performance of student subgroups is important and currently is underway. Additionally, a study investigating the longitudinal effects of exposure to the Project M$^3$ units on students who participated across all three grade levels is in progress. Further research should be conducted to investigate how participation in this curriculum might impact students' understanding of mathematics and selection of courses in middle school, high school, and beyond. Finally, although the context of this study as a curriculum development project called for a quasi-experimental design, a large-scale summative study with random assignment of students and teachers and with less professional development is warranted.

In conclusion, there is a paucity of research-based curriculum that is designed for mathematically talented students. The results of this intervention suggest that curriculum units that are concept-based, that are accelerated and enriched, and that encourage students to behave similar to practicing mathematicians contribute to students' mathematical achievement.

# References


Boix Mansilla, V., & Gardner, H. (2008). Disciplining the mind. *Educational Leadership*, *65*, 14-19.

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.

Chazan, D., & Ball, D. L. (1995). Beyond exhortations not to tell: The teacher's role in discussion-intensive mathematics classes (Craft Paper 95-2). East Lansing, MI: National Center for Research on Teacher Learning.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, *38*, 35-70.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15-42.

Confrey, J. (2006). Comparing and contrasting the National Research Council report on evaluating curricular effectiveness with the What Works Clearing House approach. *Educational Evaluation and Policy Analysis*, *28*, 195-213.

Forman, E. A. (1996). Learning mathematics as participation in classroom practice: Implications of sociocultural theory for educational reform. In L. P. Steffe, P. Nesher, P. Cobb, G. A. Goldin, & B. Greer (Eds.), *Theories of mathematical learning* (pp. 115-130). Mahwah, NJ: Lawrence Erlbaum.

Forman, E. A. (2003). A sociocultural approach to mathematics reform: Speaking, inscribing, and doing mathematics within communities of practice. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 333-352). Reston, VA: National Council of Teachers of Mathematics.

Fuson, K. C. (2004). Pre-K to grade 2 goals and standards: Achieving 21st-century mastery for all. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics: Standards in early childhood mathematics education* (pp. 105-148). Mahwah, NJ: Lawrence Erlbaum.

Gavin, M. K. (2005). Are we missing anyone? Identifying mathematically promising students. *Gifted Education Communicator*, *Fall/Winter*, 24-29.

Gavin, M. K., & Adelson, J. L. (2007). Mathematics, elementary. In Callahan, C. & Plucker, J. (Eds.), *Critical issues and practices in gifted education: What the research says.* Washington, DC: National Association for Gifted Children.

Hadamard, J. W. (1945). *Essay on the psychology of invention in the mathematical field.* Princeton, NJ: Princeton University Press.

Harcourt Brace Educational Measurement. (1997). *Naglieri Nonverbal Ability Test (NNAT)*. San Antonio, TX: Harcourt Brace.

Hiebert, J., Carpenter, T. P., Fennema, E., Fucson, K. C., Wearne, D., Murray, H., et al. (1997). *Making sense: Teaching and learning mathematics with understanding.* Portsmouth, NH: Heinemann.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Ordman, V. L., & Naylor, R. J. (2003). *The Iowa tests guide to research and development.* Itasca, IL: Riverside.

Kazemi, E. (1998). Discourse that promotes conceptual understanding. *Teaching Children Mathematics*, *4*, 410-414.

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *Journal of the Learning Sciences, 13*(1), 115-128.

Krutetskii, V. A. (1976). *The psychology of mathematical abilities in schoolchildren* (J. Teller, Trans.). Chicago: University of Chicago Press. (Original work published 1968)

Mewborn, D. S., & Huberty, P. D. (1999). Questioning your way to the Standards. *Teaching Children Mathematics, 6*, 226-227, 243-246.

Miller, R., & Mills, C. (1995). The Appalachia Model Mathematics Program for gifted students. *Roeper Review*, *18*, 138-141.

Moore, N. D., & Wood, S. S. (1988). Mathematics in elementary school: Mathematics with a gifted difference. *Roeper Review*, *10*, 231-234.

National Assessment of Educational Progress. (2008). *The nation's report card: Mathematics 2007.* Retrieve February 13, 2009 from http://nces.ed.gov/nationsreportcard/pdf/main2007/2007494.pdf

National Council of Teachers of Mathematics. (1980). *An agenda for action: Recommendations for school mathematics for the 1980s.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2007). *Curriculum research brief: Selecting the right curriculum.* Reston, VA: Author.

Pelletier, S., & Shore, B. M. (2003). The gifted learn, the novice, and the expert: Sharpening emerging views of giftedness. In D. C. Ambrose, L. Cohen, & A. J. Tannenbaum (Eds.), *Creative intelligence: Toward theoretic integration* (pp. 237-281). New York: Hampton Press.

Polya, G. (1954). *Mathematics and plausible reasoning: Induction and analogy in mathematics, Vol. 1.* Princeton, NJ: Princeton University Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage

Raudenbush, S. W., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. (2004). *HLM6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Renzulli, J. S., Leppien, J. H., & Hays, T. S. (2000). *The multiple menu model: A practical guide for developing differentiated curriculum.* Mansfield Center, CT: Creative Learning Press.

Robinson, A., Shore, B. M., & Enersen, D. L. (2007). *Best practices in gifted education: An evidence-based guide.* Waco, TX: Prufrock Press.

Robinson, A., & Stanley, T. D. (1989). Teaching to talent: Evaluating an enriched accelerated mathematics program. *Journal for the Education of the Gifted*, *12*, 253-267.

Schmidt, W. H., Wang, H. C., & McNight, C. C. (2005). Curriculum coherence: An examination of U.S. mathematics and science content standards from an international perspective. *Journal of Curriculum Studies*, *37*, 525-559.

Senk, S. L., & Thompson, D. R. (2003). (Ed.) *Standards-based school mathematics curricula: What are they? What do students learn?* Mahwah, NJ: Lawrence Erlbaum.

Sheffield, L. J. (1994). *The development of gifted and talented mathematics students and the National Council of Teachers of Mathematics Standards* (Research Monograph No. 9404). Storrs, CT: The National Research Center on the Gifted and Talented.

Sheffield, L. J. (1999). Serving the needs of the mathematically promising. In L. J. Sheffield (Ed.), *Developing mathematically promising students* (pp. 43-55). Reston, VA: National Council of Teachers of Mathematics.

Sheffield, L. J. (Chair), Bennett, J., Berriozabal, M., DeArmond, M., & Wertheimer, R. (1999). Report of the task force on the mathematically promising. In L. J. Sheffield (Ed.), *Developing mathematically promising students* (pp. 309-316). Reston, VA: National Council of Teachers of Mathematics.

Sowell, E. J. (1993). Programs for mathematically gifted students: A review of empirical research. *Gifted Child Quarterly*, *37*, 124-129.

Sowell, E. J., Bergwell, L., Zeigler, A. J., & Cartwright, R. M. (1990). Identification and description of mathematically gifted students: A review of empirical research. *Gifted Child Quarterly*, *34*, 147-154.

Sriraman, B. (2004). Gifted ninth graders' notions of proof: Investigating parallels in approaches of mathematically gifted students and professional mathematicians. *Journal for the Education of the Gifted*, *27*, 267-292.

Tieso, C. L. (2003). Ability grouping is not just tracking anymore [Electronic version]. *Roeper Review*, *26*, 29-36.

Trends in International Mathematics and Science Study. (2008). *TIMMS 2007 results.* Retrieved February 13, 2009, from http://nces.ed.gov/pubs2009/2009001.pdf

Tomlinson, C. A. (1995). Deciding to differentiate instruction in middle school: One school's journey. *Gifted Child Quarterly*, *39*, 77-87.

Tomlinson, C. A., Kaplan, S. N., Renzulli, J. S., Purcell, J., Leppien, J., & Burns, D. (2002). *The parallel curriculum: A design to develop high potential and challenge high-ability learners.* Thousand Oaks, CA: Corwin Press.

VanTassel-Baska, J., Zuo, L., Avery, L. D., & Little, C. A. (2002). A curriculum study of gifted-student learning in the language arts. *Gifted Child Quarterly*, *46*, 30-44.

Wheatley, G. H. (1983). A mathematics curriculum for the gifted and talented. *Gifted Child Quarterly*, *27*, 77-80.

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

**M. Katherine Gavin** is an associate professor at the Neag Center for Gifted Education and Talent Development at the University of Connecticut. She is currently the Principal Investigator of a National Science Foundation 5-year grant, Project M2, to develop and research advanced mathematics curriculum for students in grades K-2 as an extension of the study reported in this article. Her research interests and publications focus on identifying and developing mathematical talent, particularly in at-risk students; mathematics curriculum; and mathematical discourse.

**Tutita M. Casa**, PhD, is an assistant professor in the department of educational psychology in the Neag School of Education at the University of Connecticut. She currently is co-PI of Project M2, and is helping lead efforts to infuse purposeful mathematical writing into the curriculum. Her research interests include mathematics education, discourse, and curriculum.

**Jill L. Adelson** was a research associate at the Neag Center for Gifted Education at the University of Connecticut during the study described in this article. Currently she is an Assistant Professor in the Department of Educational & Counseling Psychology at the University of Louisville. Her research interests include the effects of gifted programming, the talent development of mathematically talented elementary students, and special issues for mathematically talented females.

**Susan Rovezzi Carroll**, PhD, is president of Words & Numbers Research, Inc which she founded in 1984. The consulting firm conducts program evaluations and research projects for educational institutions, schools and governmental agencies. She is an author of books and journal articles in the field of education.

**Linda Jensen Sheffield** is a Regents Professor Emerita of Mathematics Education and Gifted Education at Northern Kentucky University. She is chair of the NAGC Math/Science Task Force and chaired the NCTM Task Force on Mathematically Promising Students. Her main research interests are in promoting and developing mathematically promising students, and she has conducted professional development in this area across the United States as well as in sixteen other countries.