Eurasia Journal of Mathematics, Science & Technology Education, 2015, **11**(6), 1249-1264

The Impact of a Standards-based Mathematics Curriculum on Students' Mathematics Achievement: The case of Investigations in Number, Data, and Space

Ayfer Budak
Dumlupinar University, TURKEY

•Received 8 March 2015•Revised 16 April 2015 •Accepted 18 April 2015

The purpose of this research is to examine the impact of a Standards-based elementary mathematics curriculum on third grade students' mathematics performance. A total of 707 students participated in this study. Of this total, 368 students were from eight schools located within the same school district in a racially and ethnically diverse large city in the Midwest. Another 339 of the students were from four schools located in two different school districts in a middle- to high-SES, largely white, suburban area in the Northeast. Third grade students using *Investigations* curriculum outperformed in most cases and performed the same in the other cases matched comparison groups who were using a range of conventional curricula on the mathematics assessment. The *Investigations* groups did the same or better on the decontextualized, contextualized, and algebraic-reasoning constellations compared with their counterparts. The revised *Investigations* curriculum was not as effective with the low SES African-American students as it was for middle to high SES, white students. The curriculum fidelity measures did not differentiate achievement differences.

Keywords: fidelity, mathematics achievement, SES, standards-based curriculum

INTRODUCTION

Standards-based curriculum materials "embody an approach to mathematics teaching and learning that is qualitatively different from textbooks or instructional resources previously available", and in result have generated particular interest of researchers (Stein et al., 2007, p.320). Stein and her colleagues (2007) identified two reasons that increased the research to "prove" the effectiveness of these new programs: first, the restriction of the use of federal monies to those programs lacked by scientific evidence of student learning with the passage of NCLB (2002) act, and

Correspondence: Ayfer Budak,
Department of Elementary Mathematics Education, School of Education, Dumlupinar University, Kutahya, Turkey.
E-mail: ayfer.budak@dpu.edu.tr
doi: 10.12973/eurasia.2015.1377a

ISSN: 1305-8223

second, "the harsh criticism of education research and educational practices as not scientifically based led to calls for research on the effectiveness of educational programs in general, including innovative curricula such as the NSF sponsored materials " (p.320).

State of the literature

- Until recently, there had been relatively little research to extend our understanding of the effects of Standards-based textbooks on students' learning of mathematics (Ferrini-Mundy, 2003; Reys et al., 2003; Senk & Thompson, 2003).
- On most of the Standards-based curriculum studies, curriculum designers were the primary researchers which inevitably raise conflict-of-interest issues (Senk & Thompson, 2003).
- Very little research on student outcome studies stressed on teachers' fidelity to the curriculum (Putnam, 2003).

Contribution of this paper to the literature

- The purpose of this research was to examine the impact of a Standards-based elementary grades mathematics curriculum on third grade students' performance on mathematics assessment.
- In this study, the researcher had no conflict-of-interest issues with the curriculum designers.
- In this study, the impact of a Standards-based mathematics curriculum on students' mathematics achievement was investigated by taking the teachers' curriculum fidelity into account.

THE REVIEW OF LITERATURE

The effects of mathematics curricula, aligned with the NCTM (1989) Curriculum Standards and funded by the NSF, on student learning were tested by various researchers (Carter et al., 2003; Essex, 2006; Flowers, 1998; Fuson, Carroll, & Drueck, 2000; Goodrow, 1998; Huntley, Rasmussen, Villarubi, Sangtong, & Fey, 2000; McCormick, 2005; Mokros, 2000; Riordian & Noyce, 2001; Shafer, 2014, Tarr, et al., 2008). Some studies provide evidence of the positive impact of Standards-based curriculum materials on student achievement both at the elementary and middle school levels (Battistich, Allredge, & Tsuchida, 2003; Flowers, 1998; Fuson et al., 2000; Goodrow, 1998; McCormick, 2005; Reys et al., 2003; Riordian & Noyce, 2001; Shafer, 2014). In general, students in Standards-based curriculum classrooms have deeper mathematical understanding compared to the students experiencing more traditional curricula (Battistich et al., 2003; Carter et al., 2003; Flowers, 1998; Goodrow, 1998).

Research has documented that curriculum materials play a central role in mathematics instruction (Porter, 1989; Robitaille & Travers, 1992; Schmidt et al., 2001; Stein et al., 2007). In the NAEP mathematics assessments (in particular the 1992, 1996, and 2000 assessments), teachers were asked to report how frequently they use the mathematics textbook in their instruction. Results indicated that about two thirds of 4th grade students had teachers who reported that students did problems from textbooks on a daily basis (Grouws, Smith, & Sztajn, 2004). These NAEP data support the assertion that textbooks have a big influence on mathematics content taught and learned in United States classrooms. However, there are too few studies to guide researchers on how teachers' fidelity to core curriculum can be measured and is related to student outcomes (O'Donnell, 2008).

Until recently, there had been relatively little research to extend our understanding of the effects of Standards-based textbooks on students' learning of mathematics (Ferrini-Mundy, 2003; Senk & Thompson, 2003). And most of the evaluations of Standards-based mathematics programs have been limited to field studies conducted by the developers of the curricula. However, these studies provide some initial trends data concerning the impact of these programs on student achievement (Riordian & Noyce, 2001).

The studies show that students experiencing any of the following Standards-based mathematics curricula—Investigations, Math Trailblazers, Everyday Mathematics, Connected Mathematics Program (CMP), MATHematics (STEM or MT) and Number Power—do as well as or better than students exposed to other curricula on traditional measures of mathematics achievement, including computational skill. Students generally do better in interaction with other students and their teachers, and in explaining their thinking about mathematics (Carter et al., 2003); on place value and numeration (Carroll & Isaacs, 2003; Flowers, 1998; Fuson et al.,2000; Goodrow, 1998), reasoning, geometry and data, on multiplication and division involving multiples of 10, and in story or word problems (Carroll & Isaacs, 2003; Fuson et al., 2000;Mokros et al, 1994); on addition and subtraction with regrouping (Carroll & Isaacs, 2003; Goodrow, 1998), and on writing number sentence for a number (Goodrow, 1998), on open ended and problem solving mathematics tests (Post et al, 2008). In general, they have deeper mathematical understanding compared to the students experiencing more traditional curricula ((Battistich et al., 2003; Carter et al., 2003; Flowers, 1998; Goodrow, 1998).

With respect to the issues of equity in Standards-based curricula, the studies do not encouraging results: McCormick (2005) found that the revised Investigations curriculum seems to favor white, and high SES students. The results of a more current study on Everday Math at three rural sites showed that the constructivist K-6 elementary mathematics curriculum did not lead to higher levels in math achievement when compared with the traditional curricula (Grady, Watkins and Montalvo, 2012).

Four studies were conducted aiming to examine the effects of the Investigations in Number, Data, and Space on children understanding of number and number operations. Mokros (2003) summarized previous studies of Investigations. Investigations students did as well on mastery of basic facts as students using other curricula (Flowers, 1998; Goodrow, 1998; Mokros, Berle-carman, Rubin, & Wright, 1994) with no differences between groups with respect to accuracy. Investigations students did as well as or better than students using other curricula on calculation problems (Flowers, 1998; Goodrow, 1998; Mokros et al., 1994). Investigations students achieved greater accuracy than students using comparison curricula on word problems and on more complex calculations, for example, on word problems in which the use of a particular arithmetic operation was not specified (Mokros et al., 1994) or problems involving proportional reasoning (Flowers, 1998). Implementation of the Investigations curriculum had the greatest impact when students were encouraged to develop their own strategies and when teachers did not combine Investigations with a more traditional approach to teaching algorithms (Flowers, 1998; Goodrow, 1998; Mokros, Berle-Carman, Rubin, & O'Neil, 1996).

The Investigations curriculum was revised based on the findings of the studies discussed above. The purpose of this researchwas to examine the impact of a Standards-based elementary grades mathematics curriculum on third grade students' performance on mathematics assessment. The curriculum of interest is revised Investigations in Number, Data, and Space developed by TERC under a grant from NSF. To pursue this purpose, the following four questions were explored. What differences exist between the mathematics achievement:

a) of the students in high-fidelity Inv. classrooms and those in the high-fidelity non-Inv. classrooms? (between curricula comparisons)
b) of the Inv. students in high-fidelity classrooms to those in the low-fidelity classrooms (within Inv. comparisons)?
c) of the non- Inv. students in high-fidelity classrooms to those in the low-fidelity classrooms (within non- Inv.)?
d) of the low-fidelity Inv. classrooms and those in the low-fidelity non-Inv. classrooms? (between curricula comparisons)

METHOD

Participants

A total of 707 students participated in this study. Of this total, 368 students were from eight schools located within the same school district in a racially and ethnically diverse large city in the Midwest of USA. In the participating schools, about 69 % of the revised Investigations students and 63 % of non-Investigations students were eligible for free or reduced lunch. The majority of both Investigations and non-Investigations students were composed of non-white students (see Figure 1). To protect the privacy of participants, I refer to this city as "Metroville" throughout this paper.

Another 339 of the students were from four schools located in two different school districts in a middle- to high-SES, largely white, suburban area in the Northeast of USA. The majority of the student populations were composed of white students for both Investigations and non-Investigations groups (See Figure 2). Five percent of the Investigations (n = 118) and again only five percent of the non-Investigations (n =221) students were eligible for free or reduced lunch.

A total of 47 third-grade teachers participated in this study. Of these 47 teachers, 26 were from Metroville and sixteen of whom reported that they used the revised Investigations curriculum as the primary text. The other ten teachers reported that

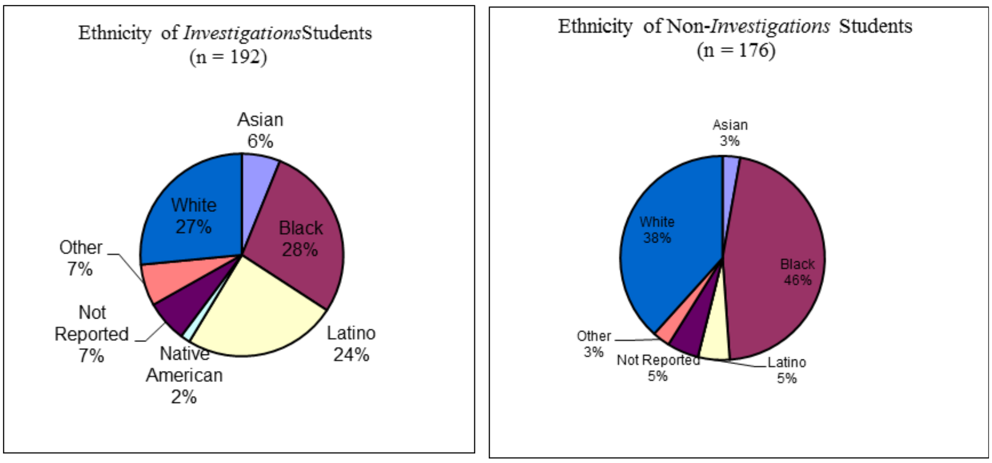

Figure 1. Ethnicity of *Investigations* and non-*Investigations* students in the Metroville school district

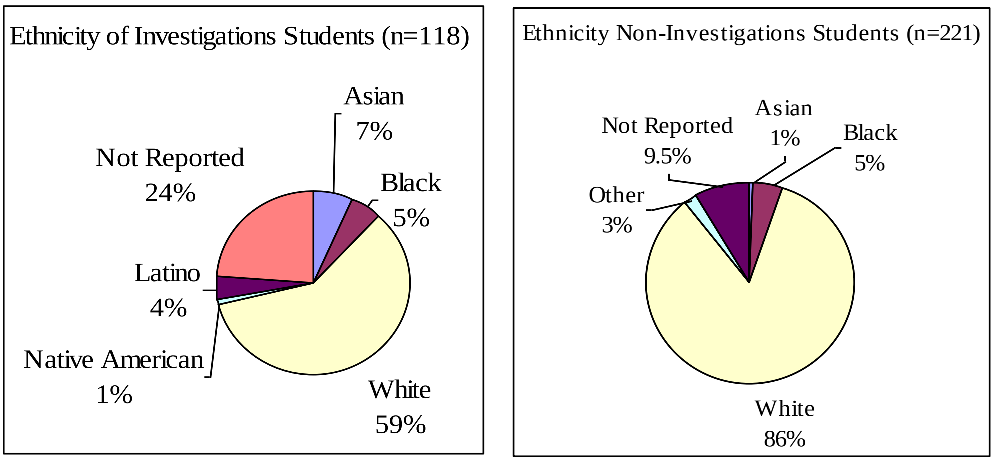

Figure 2. Ethnicity of *Investigations* students in the Eastville school district and non-*Investigations* students in the Northville school district

they used one of the comparison curricula as their primary text. Of the 21 teachers from Northville and Eastville[1] districts, seven reported that they used the revised Investigations curriculum as their primary text, and the remaining fourteen reported that they used one of the comparison curricula as their primary text.

The participating teachers from the three districts had a range of teaching experience anywhere from 1 to 40 years[2] at the elementary school level. The average number of years the Investigations teachers reported to have taught was eight, and the average number of years the comparison group teachers reported to have taught was fourteen. Twelve out of 24 non-Investigations teachers and 14 out of 23 Investigations teachers reported to have participated in curriculum-specific professional development (mainly inservice workshops) at the school level or both at the school and the district level.

Curricula

Investigations in Number, Data, and Space aims to accomplish six major goals: (1) to provide meaningful problems requiring students to think mathematically, encouraging the use of different strategies by students with different learning styles; (2) to develop powerful mathematical thinking, explanation, justification and demonstration; (3) to encourage sustained thinking; (4) to provide coherent and in-depth mathematical content; (5) to support teacher learning; and (6) to connect all students- students with a range of abilities, ethnicity, or language- with mathematics (Mokros, 2003).

The comparison groups, non-Investigations students used commercially developed conventional mathematics curricula: In Northville school district, the comparison schools were using Silver Burdett Ginn Mathematics. The comparison schools in Metroville school districts were using either one of the three conventional curricula: Houghton Mifflin, McGraw-Hill Math, Connecting Math Concepts. As most Standards-based curricula, conventional curricula also intend for students learn concepts, skills, applications, problem solving and efficient procedures. However, Stein et al., (2007) noted that they differ in several ways: Conventional curricula tend to (a)"rely on direct explication of the to-be-learned material as well as careful sequencing and accumulation of lower-level skills before presenting students with the opportunity to engage in higher-order thinking, reasoning and problem solving with those skills" (p.331), (b) deepen students' understanding by attaining eventual mastery over time as the topic is revisited, (c) focus on practicing procedures, and (d) discourage the use of calculators on the development of computational skills.

Instruments and data collection

Two types of instruments were used to collect data for this study: Math Assessment and teacher curriculum logs.

Mathematics assessment

The IU Curriculum Evaluation Research Team administered the third grade mathematics assessment they developed in response to TERC's prioritized benchmarks. The math test was administered in the fall and in the late spring of the school year. The math assessment contains eight problem contexts that include 23 individual items and it takes students about 50-60 minutes to complete. The assessment includes items on number sense and operation and algebraic reasoning

[1] To protect the privacy of the participants from these school districts, I refer to one school district as "Eastville" and the other district as "Northville."
[2] Two *Investigations* teachers and one non-*Investigations* teacher in Metroville failed to report the total number of years in teaching.

and was designed to capture the growth in students' learning over a school year. The assessment emphasizes problem-solving contexts and authentic tasks over symbolic computation (Kehle et al, 2004). The main goal of this assessment was to examine students' problem-solving ability, use of computational strategies, mathematical reasoning and number sense.

In addition to looking at the overall scores on the Math assessment, students' achievements on constellations or groups of problems: Decontextualized computation, contextualized computation, and algebraic reasoning constellations were also investigated. The decontextualized-computation constellation included four problems: two multi-digit addition, one multiplication, and one subtraction problem. These problems were purely computational and without a context. Within this constellation, there was a total of eight possible points. The contextualized-computation constellation characterizes the computational problems that are set within a context; they may have also been called story or word problems. This constellation included six problems, and within this constellation there was a total of 14 possible points. The algebraic-reasoning constellation includes problems that can "capture students' ability to describe, extend, represent, and make generalizations about quantitative relationships and patterns" (McCormick, 2005, p.58). Within this constellation, there was a total of seven possible points.

Curriculum logbook

Curriculum logbooks were used to describe the curriculum teachers implemented in the classroom. Beyond knowing what primary curriculum a teacher was using, the data from this instrument indicate the fraction of the curriculum used, the content emphasized, teaching methods used, the specific focus of the students, materials and resources used, and assessment techniques employed for each of the ten lessons. Teachers also reported if the content studied varied by groups of students, the reason why these groups received different instruction, and any difficulties students had associated with various topics.

Teachers were given. Each logbook contains written instructions to fill out the logs, data on 10 consecutive days of mathematics instruction and was adapted from an instrument used by Romberg and Shafer (2003). These logbooks were collected from the third-grade teachers twice during the school year.

Procedures

Validity of the math assessment

Content validity is the most important type of validity for achievement tests (Ravid, 1994). To increase the validity of the third-grade mathematics assessment instrument, items were borrowed whenever possible from existing instruments (Kehle et al, 2004). Teachers reviewed the tasks for appropriateness and validity. An advisory board provided feedback on the validity and curriculum neutrality of the items. The instrument was field tested at non-Investigations sites and non-Investigations teachers checked to see if any of the items were biased in favor of or against either group of students (Kehle et al., 2004). The field testing and the review of the instruments by non-Investigations teachers provided strong evidence that the instrument was curriculum-neutral (McCormick, 2005).

Developing the coding rubric for the math assessment

A coding rubric was developed by the team of researchers. The rubric was designed specifically for individual tasks. The following factors were considered when developing the rubric: accuracy of the solution, common error patterns, level of understanding of the context, and the strategies or approaches used in solving the problems. The rubric was revised several times on the basis of the coders' comments and discrepancies and finalized.

The coders were kept blind to the students' identities, their schools and districts, which semester the test was administered, and the curriculum being studied. First, three researchers coded different sets of assessments to catch the major discrepancies and met several times to resolve the discrepancies that came across during these reliability attempts. After adjusting for discrepancies, the overall estimates intercoder reliability (Vierra, Pollock, & Golez, 1998) was 91% with the reliability estimates for individual items ranging from 89 - 93%.

Scoring the mathematics assessment

The scoring rubric was designed using a holistic technique. As suggested by Perlman (2003), each point on the scale was clearly labeled and defined. The scoring rubric consists of five categories: low, medium-low, medium, medium-high, and high. High represents a correct answer with a reasonable work or justification; medium generally represents a correct answer with no or little reasonable work or justification or an incorrect answer (e.g., computational error) with a work that demonstrates child's correct mathematical understanding of the problem; and low represents incorrect answer and the work demonstrates very little or no mathematical understanding of the problem. The medium-low and medium-high categories were used to score the items with multiple levels These five categories were used to score each item except for the purely computational problems and correct/incorrect type items. Those types of items were scored as either high or low. Between 0-2 points were assigned for each of the five categories: 0 point (for low), 0.5 point (for medium-low), 1 point (for medium), 1.5 point for (medium- high), and 2 point (for high). Students with some sophisticated work could receive 1 extra point on the three problems: Boxes of Bags of Rocks, Ann's Towers, and the third part of the Helpful Computing. Including the sophistication points, a total of 45 points were possible for the assessment (7 points for algebraic reasoning and 38 for number and operations). The inter-problem estimate (Cronbach's alpha) for the fall assessment was .82, and .85 for the spring assessment. Estimates of split-half reliability values range from .76 to .79.

Fidelity to the curriculum

Curriculum logs were analyzed to check the teachers' fidelity to the primary curriculum. The logbooks provide information regarding how often the primary text was used during the 20-day school period (10 days from each semester). In logbooks, teachers were asked to check all materials they used for the day's lesson or for that night's homework. Teachers were allowed to put more than one check mark whenever it is appropriate.

All logbooks from third grade teachers were coded using the following system: if teachers reported using either the primary textbook and/or supplemental materials from textbook publisher not in student or teacher editions, they were assigned the high code; if teachers reported using either teacher-designed (by you) materials, and/or worksheets or materials from other sources and/or any other materials, they were assigned the low code; and any log that could not be labeled either high or low was coded as medium.

Frequencies were calculated for each of the three codes out of the number of days that teachers filled out the logs. Finally, weighted percentages were calculated using the following formula:

Zero-point weight to the code low; 1-point weight to the code medium and 2-point weight was assigned to the code high. After calculating the weighted percentages per teacher, the quartiles were calculated to categorize teachers' use of textbooks. Thirteen teachers' weighted percentages were at or above the 3rd quartile and those teachers were labeled as primary text users with *high fidelity (92%)*. Thirteen teachers' weighted percentages were at or below the 1st quartile, and those teachers were labeled as *low- fidelity (72%)* primary text users.

Data analysis

Hake (1999) pointed out that the normalized gain was well established as a rational method in analyzing pre-test, post-test scores in the physics and astronomy research, yet, it is not well recognized in educational research. Analysis of Covariance (ANCOVA) and t-test on the gain scores are the two main approaches partialling out the initial differences (Wright, 2005). All the analysis on the mathematics assessment focuses on normalized gain[3] scores. Simply, it is the percentage of the possible gain that was in fact realized by a student. The normalized gain highly correlates (r = 0.96, p < .001) with the actual gain, but it better represents the improvement of students across a broad range of situations.

A pre- and posttest, quasi-experimental design was used in this study. Computer software, SPSS, was used in analyzing the data. The American Psychological Association (APA) Board of Scientific Inference (1998) has urged researchers to accompany their p values with an estimate of both the direction and size of an effect. Effect size estimates provide first step toward evaluating the practical importance of a finding (McCartney & Rosenthal, 2000). Hedge's g, an effect-size statistic, was used in this study (Kirk, 1999). Like Cohen's d, Hedge's g is used in assessing the relations via comparison of group means and can be interpreted in the same manner. Specifically, for Hedge's g, these guides are: a g of .20 is small, .50 moderate, and .80 large. In this study, the effect size was reported for each group's normalized gain scores that were significantly different.

RESULTS

The answers to the four research questions were discussed separately.

High-fidelity inv. vs. high-fidelity non-inv.

This portion of the results includes a comparison of the students in high-fidelity Inv. and non-Inv. Classrooms. A total of six Inv. classrooms (four from Metroville and two from Eastville), and a total of five non-Inv (three from Metroville and two from Northville) classrooms were part of the analysis in this section. These eleven teachers were highly dependent on their mathematics curriculum.

Between curricula

An independent samples t-test was conducted to evaluate the hypothesis that there is no significant difference in mathematics achievement between the students of Inv. teachers and the students of non-Inv teachers with high fidelity to their curricula.

Table 1 shows that there was a significant difference between the two groups' Mean Gain scores[4] (MGs) in favor of high-fidelity Inv. students (t (165.07) = 2.78, p <. 01) and the effect size g of 0.43 indicates a medium-low effect. Likewise, the Mean Normalized Gain scores (MNGs) of the two groups were significantly different in

[3] Normalized Gain = (Post-test – Pre-test) / (Maximum on test - Pre-test)
[4]*Note.* The Mean Gain scores (MGs) refer to the mean growth that was made on the Math assessment from Fall to Spring.

Table 1. The overall mean gains and mean normalized gains on the math assessment delineated by curriculum and fidelity

	High-Fid. Inv. (n= 92)		High-Fid. Non-Inv. (n= 80)		
Scores	**M**	**SD**	**M**	**SD**	**t**
Gain	11.12	8.16	8.13	5.94	2.78**
N-gain	0.34	0.25	0.25	0.18	2.64**

*Note. **p<.01*

Table 2. The high-fidelity ınv. students' overall mean gains and mean normalized gains on the math assessment delineated by school district

	High-Fid. Inv.				High-Fid. Non-Inv.				Pairwise Comparisons			
	Metro. (n= 58)		East. (n= 34)		Metro. (n= 38)		North. (n= 42)					
Scores	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	t_{Inv}	t_{Non}	t_{Metro}	$t_{East-North}$
Gain	7.81	5.73	16.77	8.65	7.45	5.78	8.74	6.08	-5.38***a	-0.92	0.30	4.57***c
N-Gain	0.20	0.15	0.57	0.22	0.21	0.16	0.28	0.19	-8.56***b	-1.72	-0.34	6.03***d

*Note. [a,b,c,d] The effect size g indicates a high-large effect. ***p<.001*

favor of the high-fidelity Inv. students ($t(164.37) = 2.64$, $p < .01$). The effect size g of 0.40 indicates a medium-low effect.

The high-fidelity Inv. group had higher MNGs on all of the three constellation problems compared to the high-fidelity non-Inv. group. However, the gain differences were not significant.

When comparing the high-fidelity Inv. students' MGs and MNGs on the Math assessment, Table 2 shows that Inv. students in the Eastville district have significantly ($p < .001$) higher means compared to those in the Metroville district. A similar analysis of the gains made on the non-Inv. students indicates that there were not significant differences in the gains made over the school year or MNGs of the Metroville and Northville districts' non-Inv. students.

Within school districts

To take a different look at the achievement scores of high-fidelity Inv. and non-Inv. groups, I compared the scores of students within both districts, Metroville and Northville-Eastville.

When looking at the mean normalized gain scores of the Inv. in the Eastville district and non-Inv groups in the Northville district on the three constellation problems, the Inv. students' MNGs (M = 0.53, SD = 0.41) on the contextualized computation constellation were significantly higher ($t(56.75)= 2.93$, $p < .005$) higher than that of their non-Inv. counterparts (M=0.28, SD=0.28). The effect size g of 0.68 indicates a medium-high effect. This was the only significant difference between the two groups on the constellations problems.

High- vs. low-fidelity inv.

A total of five low-fidelity and six high-fidelity Inv. classrooms were included in this analysis.

Within a curriculum

An independent samples t-test was conducted to evaluate the hypothesis that there is no significant difference in mathematics achievement between the students in high-fidelity Inv. classrooms and low-fidelity Inv. classrooms. As it is presented in Table 3, the MGs and the MNGs that were made over the school year did not differ

Table 3. The inv. students' overall mean gains and mean normalized gains on the math assessment delineated by fidelity

	High-Fid. Inv. (n=92)		Low-Fid. Inv. (n=109)		
Scores	**M**	**SD**	**M**	**SD**	**t**
Gain	11.12	8.16	12.43	7.10	-1.21
N-gain	0.34	0.25	0.34	0.20	0.06

Table 4. The metroville district inv. students' overall mean gains and mean normalized gains on the math assessment delineated by district and fidelity

	Metroville				Eastville				Pairwise Comparisons	
	High-Fid. Inv. (n= 58)		Low-Fid. Inv. (n= 89)		High-Fid. Inv. (n= 34)		Low-Fid. Inv. (n= 20)			
Scores	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	$t_{Metroville}$	$t_{Eastville}$
Gain	7.81	5.73	10.89	6.61	16.76	8.65	19.25	4.92	-2.91***	-1.35
N-Gain	0.20	0.15	0.28	0.18	0.57	0.22	0.57	0.12	-2.81**	-0.14

*Note. **p < .01, ***p < .005*

Table 5. The non-inv. students' mean gains and mean normalized gains on the math assessment delineated by fidelity

	High-Fid. non-Inv. (n = 80)		Low-Fid. non-Inv. (n = 83)		
Scores	**M**	**SD**	**M**	**SD**	**t**
Gain	8.13	5.94	10.11	7.25	-1.91
N-gain	0.25	0.18	0.30	0.22	-1.76

significantly. In addition, no significant differences were found between the two groups' MGs on the three constellations.

Within districts

Table 4 shows MGs and MNGs of both groups on the Math assessment. The groups were defined by the fidelity to the curriculum and the school district.

Within the Metroville school district, mean normalized gain scores (see Table 4) were significantly greater (p <.005, p <.01, respectively) than that of their high-fid. Inv. counterparts. To gain a better understanding of this data, the scores of these two groups on the three constellations were compared. No difference was found between the groups'performane on three constellations.

When looking at the results within Eastville school district Table 4 shows that there were no significant differences on the MGs and MNGs over the school year for the high- and low-fidelity Inv. students. Similar analyses were run on the three constellations of problems. When comparing the high-fidelity Inv. and low-fidelity Inv. groups' MNGs on the three constellations within the Eastville school district, there was no significant difference between the two groups' mean scores.

High- vs. low-fidelity non-inv.

In this section, the analyses of MGs and MNGs of low- and high-fid. non-Inv students are presented. The groups were defined by fidelity and school districts.

An independent samples t-test was conducted to evaluate the hypothesis that there is no significant difference between the students in high- and low-fidelity non-Inv. classrooms on mathematics achievement (see Table 5).

Table 6. The metroville non-inv. students' mean gains and mean normalized gains on the math assessment delineated by district and fidelity

	Metroville				Northville				Pairwise Comparisons	
	High-Fid. Non-Inv (n= 38)		Low-Fid. Non-Inv (n= 31)		High-Fid. Non-Inv (n= 42)		Low-Fid. Non-Inv (n= 52)			
Scores	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	$t_{Metroville}$	$t_{Northville}$
Gain	7.45	5.78	6.36	6.55	8.74	6.08	12.35	6.76	0.74	-2.69**[a]
NGain	0.21	0.16	0.19	0.21	0.28	0.19	0.37	0.19	0.44	-2.23*[b]

*Note. [a] The effect size g of 0.56 indicates a medium effect. [b] The effect size g of 0.46 indicates a medium-low effect. *p<.05, and **p<.01.*

Table 7. The overall mean gains and mean normalized gains on the math assessment delineated by curriculum and fidelity

	Low-Fid. Inv. (n= 109)		Low-Fid. Non-Inv. (n= 83)		
Scores	**M**	**SD**	**M**	**SD**	**t**
Gain	12.43	7.10	10.11	7.25	2.22*
N-gain	0.34	0.20	0.30	0.22	1.01

*Note. *p<.05*

Within districts

Table 6 shows MGs and MNGs of the non-Inv. students in the low- and high-fidelity classrooms on the Math assessment. The groups were defined by the fidelity to the curriculum and the school district.

Within the Metroville school district, Table 6 shows that the students' MGs over the school year and MNGs did not differ significantly. The two groups' mean scores on the three constellations were also compared and no significant difference was found on none of the constellations.

Within Northville school district, when looking at the gains made over the school year and MNGs, the two groups' mean scores differed significantly ($p < .01$, $p < .05$, respectively) in favor of the low-fidelity non-Inv. students (see Table 6). No significant difference was found on any of the constellations.

Low-fidelity inv. vs. low-fidelity non-inv.

This section includes the analysis of the low-fidelity students' achievements on the math assessment and on three constellation problems.

Between curricula

This section presents the results of the analysis between the low-fidelity Inv. and low-fidelity non-Inv.

However, Table 7 shows that the low-fidelity Inv. group had a significantly greater ($p < .05$) performance gain than their non-Inv. counterparts. The effect size g of 0.32 indicates a low effect. The two groups' MNGs did not differ significantly. A similar analysis was done on the three constellation problems, and no significant differences were found on any of the constellations.

Table 8 shows the students' MGs and MNGs on the math assessment. The students were placed on these groups based on their mathematics teacher low-fidelity to their primary mathematics curriculum.

Within curriculum

This part of the analysis compares the mean scores of the Inv. and non-Inv. students in low-fidelity classrooms. Table 8 shows the MGs over the school year and

Table 8. The Fall and Spring Mean Scores of the Inv. Students in the Low Fidelity Classrooms on the Math Assessment Delineated by District

	Low-Fid. Inv.				Low-Fid. Non-Inv.				Pairwise Comparisons			
	Metroville (n= 89)		Eastville (n= 20)		Metroville (n= 31)		Northville (n= 52)					
Scores	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	t_{Inv}	t_{Non}	t_{Metro}	$t_{East-North}$
Gain	10.89	6.61	19.25	4.92	6.36	6.55	12.35	6.76	-5.33***a	-3.95***c	3.30***e	4.16***g
N-Gain	0.28	0.18	0.57	0.12	0.19	0.21	0.37	0.19	-7.07***b	-3.93***d	2.28*f	4.44***h

Note. [a] *The effect size g of 1.32 indicates a large effect.* [b] *The effect size g of 1.75 indicates a large effect.* [c] *The effect size g of 0.90 indicates a large effect.* [d] *The effect size g of 0.89 indicates a large effect.* [e] *The effect size g of 0.69 indicates a medium effect.* [f] *The effect size g of 0.48 indicates a medium-low effect.* [g] *The effect size g of 1.09 indicates a large effect.* [h] *The effect size g of 1.17 indicates a large effect.*
**p < .05, ***p < .001*

MNGs. Again, the Inv. students in the Eastville district had significantly (p < .001) higher MGs and MNGs than their counterparts in the Metroville district. When comparing the MNGs of the Inv. students in Metroville and Eastville on the three constellations, the two groups' mean scores differed significantly on all of the three constellations (t (107) = -2.69, p < .01, and with a g of 0.67 on the decontextualized constellation; (t (107) = -4.60, p < .001, and with a g of 1.14 on the contextualized constellation; and t (107) = -6.74, p < .001, and with a g of 1.67 algebraic-reasoning constellation) in favor of the low-fidelity Inv. students in Eastville.

Similarly, Table 8 shows that the low-fidelity non-Inv. students in Northville had significantly (p < .001) higher mean scores than the low-fidelity non-Inv. in Metroville. When looking at the three constellations, there was a significant difference between the two groups MNGs (t (81) = -2.52, p < .05, and with a g of 0.57) on the contextualized, and (t (81) = -2.33, p < .05, and with a g of 0.53) on the algebraic-reasoning constellations in favor of low-fidelity non-Inv. students in Northville school district.

Within districts

Within Metroville school district, the low-fidelity Inv. students had significantly greater MGs and MNGs over the school year (p < 001, and p < .05 respectively) on the mathematics assessment than the students in low-fidelity non-Inv. classrooms (see Table 8). When comparing the two groups' mean scores on the three constellations, there was no significant difference on any of the three constellations.

Within the Eastville and Northville school districts, the Inv. students in Eastville had significantly greater MGs and MNGs (p < .001) than the non-Inv. students in Northville (see Table 8). For the Eastville and Northville districts, the MNGs of the low-fidelity Inv. students in Eastville and non-Inv. students in Northville were compared on the three constellations. The results reveal that the low-fidelity Inv. students in Eastville performed significantly better on two of the constellations (t (70) = 2.43, p < .05, and with a g of 0.64 on the contextualized constellation, and t (70) = 2.83, p < .01, and with a g of 0.75 on the algebraic-reasoning constellation) than the non-Inv. students in Northville.

DISCUSSION AND CONCLUSION

The impact of mathematics curricula on mathematics achievement in the classrooms where teachers had high- or low- curriculum fidelity was investigated. The findings suggest that the Investigations students performed the same or significantly outperformed the non-Investigations students on the mathematics assessment irrespective of teachers' fidelity to the curriculum and the school districts. In addition, although not significant, in most cases the Investigations students had higher gain scores on the three constellation problems: decontextualized, contextualized, and algebraic reasoning compared to the non-

Investigations students; and in some cases significant higher gains on one or two of the constellations or all three. These findings are consistent with much of the findings of previous literature on the impact of Standards-based elementary mathematics curricula on student achievement (Flowers, 1998; Goodrow, 1998; Mokros et al., 1994). This interesting finding may mislead the reader with regard to the importance of fidelity to the curriculum. The curriculum fidelity measures used in this study did not differentiate achievement differences in the classrooms with high and low fidelity to the curriculum. Certainly, one could not claim that teachers' fidelity to the curriculum is not an important variable in curriculum research such as this one. The main reason behind these findings might be the fact that the fidelity measures included only teachers' own reports of their use of curriculum and did not include classroom observations.

The results show that the middle to high SES white Investigations students had significantly higher scores on the mathematics assessment than the low SES African American, Hispanic or Latino Investigations students irrespective of teachers' fidelity to the curriculum. This finding is consistent with the NAEP 2000 mathematics results that the average scale score of the students who were white was considerably higher than that of the African-American and Latino students at each grade level (4th, 8th, and 12th) (Strutchens, Lubienski, McGraw, & Westbrook, 2004). The white students performed significantly better than African American and Latino students within each content area on the assessment (e.g., number sense, properties, operations, and algebra). Similarly, the NAEP 2000 mathematics assessment results indicate that the high-SES students performed significantly had higher average scale score than the low-SES students. These findings support McCormick (2005)'s finding that "the revised Investigations curriculum produced more inequitable levels of achievement" for the low SES and African-American, and Latino students. (p. vii).

At school district with the majority of low SES, African American, Hispanic or Latino students, no significant difference was found between the high-fidelity Investigations and non-Investigations groups on the mathematics performance, but when looking at the results in low-fidelity Inv. and non-Inv. classrooms, there was a significant difference between the two groups, in favor of the Investigations students. This result suggests that the revised Investigations curriculum did not make any difference on the achievement of disadvantaged students. However, the students whose teachers reported using extra materials or some other curricula in their instruction as well as the Investigations curriculum performed better on the mathematics assessment compared with the students whose teachers reported using additional materials or curriculum other than the primary conventional textbook. Clearly, neither of the curricula was the factor affecting the student achievement. One explanation to the high achievement of the students in low-fidelity Investigations classrooms could be that the teachers in these classrooms might be using some other curricula or teaching methods that were very effective for this group of student population. Students benefit more from the Standards-based curricula when it's coupled with a Standards-based learning environment (Tarr et al., 2008).

Suggestions for future research

This study raises a number of questions that suggest a need for further study. Teachers' fidelity to the curriculum should be investigated more deeply, and should be reported for all studies examining the effects of one program on students' learning. Fidelity to the curriculum is an important variable to interpret the results of comparative studies such as this one. However, most studies on the impact of Standards-based curricula on student achievement have left out this important variable. I suggest that the fidelity measures should include observational data and/or include more data regarding the extent too which the curriculum was carried out in the classroom as intended by the curriculum developers. Also, It would be helpful to develop curriculum logs that can reveal both teachers' implementation of and fidelity to the curriculum. To this end, teachers should be provided with clearer explanations and descriptions of terms that might be subject to different interpretations. For instance, in this study, some teachers interpreted some of the items differently for other teachers and some teachers recorded their answers in the wrong place. In order to overcome this, teachers should be trained with regard to filling out the curriculum logs and should be provided a sample curriculum log or a log could be filled out after observing a lesson of a colleague or a project assistant, etc.

Finally, two suggestions come to mind that are more specific to the Investigations curriculum, but in fact might apply to other Standards-based mathematics curricula as well. First, previous research, as well as, this present study suggests that the Investigations curriculum is not as effective with low SES African-American student as it is for white students. An in depth examination is needed of the content and presentation of Investigations curriculum aimed at uncovering any possible racial or ethnic biases. Second, more study is needed to examine of the performance of low achieving low SES Investigations classrooms over time to get a better sense of the interaction between the curriculum and this particular group of students.

Third grade students using Investigations outperformed on most cases and performed the same in some cases matched comparison groups who were using a range of conventional curricula on the mathematics assessment. The Investigations groups did the same or better on the decontextualized, contextualized, and algebraic-reasoning constellations compared with their counterparts. While the gain in student performance was greater with middle to high SES, white students who were using Investigations curriculum, low SES, African American and Hispanic students benefitted more from the Standards-based teaching practices than the use of Investigations curriculum.

Writing about their study on the impact of Standards-based curricula on the middle school students' mathematics performance, Reys et al. (2003, p.89) concluded that "just as an artist's picture takes shape with each stroke of the brush, the critical development of a strong research base to investigate the effect of Standards-based curriculum materials on student learning will not be established by any single study." I offer this study as one more "stroke of the brush" to the emerging picture we have about the effect Standards-based mathematics curricula have on students' learning.

REFERENCES

Battistich, V., Alldredge, S., & Tsuchida, I. (2003). *Number Power*: An elementary school program to enhance students' mathematical and social development. In S. Senk & D. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 133-160). Mahwah, NJ: Lawrence Erlbaum Associates.

Carroll, W. M., & Isaacs, A. (2003). Achievement of students using the University of Chicago School Mathematics Project's *Everyday Mathematics*. In S. Senk & D. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 45-78). Mahwah, NJ: Lawrence Erlbaum Associates.

Carter, A., Beissinger, J. S., Cirulis, A., Gartzman, M., Kelso, C., & Wagreich, P. (2003). Student learning and achievement with *Math Trailblazers*. In S. Senk & D. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 45-78). Mahwah, NJ: Lawrence Erlbaum Associates.

Essex, N. K. (2006). *Looking for gender differences in the mathematical work of elementary students*. Unpublished doctoral dissertation, Indiana University, Bloomington.

Ferrini-Mundy, J. (2003). What does it mean to be Standards-based? Issues in conceptualizing, measuring, and studying alignment with standards. In F. K. Lester & J. Ferrini-Mundy (Eds.), *Proceeding of the NCTM research catalyst conference*. Reston, VA: National Council of Teachers of Mathematics.

Flowers, J. (1998). *A study of proportional reasoning as it relates to the development of multiplication concepts.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Fuson, K., Carroll, W. M., & Drueck, J. V. (2000). Achievement results for second and third graders using the standards-based curriculum *Everyday Mathematics*. *Journal for Research in Mathematics Education*, *31*(3), 277-295. doi:10.2307/749810

Grouws, D. A., Smith, M.S., & Sztajn, P. (2004). The preparation and teaching practices of United States Mathematics Teachers: Grades 4 and 8. In Kloosterman, P., & Lester, F. K. (Eds.) *Results and Interpretations of the 1990-2000 Mathematics Assessments of National Assessment of Educational Progress.* Reston, VA: The National Council of Teachers of Mathematics.

Hake, R. R. (1999). Analyzing change/gain scores. Retrieved March 1, 2015, from http://www.physics.indiana.edu/~sdi/AnalyzingChange-Gain.pdf

Huntley, M. A., Rasmussen, C. L., Villarubi, R. S., Sangtong, J., & Fey, J. T. (2000). Effects of standards-based mathematics education: A study of the Core-Plus mathematics Project algebra and functions strand. *Journal for Research in Mathematics Education, 31*(3), 328-361. doi:10.2307/749810

Kehle, P. E., Lambdin, D.V. Essex, N. K., & McCormick, K. K. (2004). *How will we know what they learned? Designing a longitudinal study of elementary mathematics curriculum.* Paper presented at the meeting of the American Education Research Association, San Diego, CA.

Kirk, R. E. (1999). *Statistics-An Introduction.* Orlando: Harcourt Brace & Company.

McCartney, K. & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development, 71*, 173-180. doi:10.1111/1467-8624.00131

McCormick, K. (2005). *Examining the relationship between a standards-based elementary mathematics curriculum and issues of equity*. Unpublished doctoral dissertation, Indiana University, Bloomington.

Mokros, J. (2000). *The Investigations curriculum and children's understanding of whole number operations.* Cambridge, MA: TERC. Available from the author, http://www.terc.edu/investigations/impact/html/impact-map.html

Mokros, J. (2003). Learning to reason numerically: The impact of *Investigations.* In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 471-488). Mahwah, NJ: Lawrence Erlbaum Associates.

Mokros, J., Berle-Carman, M., Rubin, A., & O'Neil, K. (1996, April). *Learning operations: Invented strategies that work.* Paper presented at the meeting of the American Educational Research Association, New York.

Mokros, J., Berle-carman, M., Rubin, A., & Wright, T. (1994, December). *Full year pilot grades 3 and 4: Investigations in number data and space*. Cambridge, MA: TERC. Available from the author, http://www. terc.edu/investigations/impact/html/impact-map.html

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33-84. doi:10.3102/0034654307313793

Perlman, C. C. (2003). *Performance assessment: Designing appropriate performance tasks and scoring rubrics.* (ERIC Document Reproduction Service No. ED 480070).

Porter, A. (1989). A curriculum out of balance: The case of elementary school mathematics. *Educational Researcher, 18*(5), 9-15. doi:10.3102/0013189x018005009

Post, T. R., Harwell, M. R., Davis, J. D., Maeda, Y., Cutler, A., Andersen, E.,& Norman, K. W. (2008). Standards-based mathematics curricula and middle-grades students' performance on standardized achievement tests. *Journal for Research in Mathematics Education, 39*(2), 184-212.

Putnam, R. T. (2003). Commentary on four elementary mathematics curricula. In S.L. Senk & D.R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 161–178). Mahwah, NJ: Erlbaum.

Ravid, R. (1994). *Practical Statistics for Educators*. Lanham: University Press of America.

Reys, B. J., Reys,R. E., Hope, J. A. (1993). Mental computation: A snapshot of second, fifth, and seventh grade student performance. *School Science and Mathematics, 93,* 306-315.

Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of Standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education, 34*, 74-95 doi:10.2307/30034700

Riordian, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32*, 368-398. doi:10.2307/749700

Robitaille, D. F., & Travers, K. J. (1992). International studies of achievement in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 687-709). New York: Macmillan.

Romberg, T. A., & Shafer, M. (2003). Mathematics in Context (MiC): Preliminary evidence about student outcomes. In S. Senk & D. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* Mahwah, NJ: Lawrence Erlbaum Associates.

Schmidt, W. K., McKnight, C.C., Houang, R. T., Wang, H.C., Wiley, D. E., Cogan, L. S., et al. (2001). *Why schools matter: A cross-national comparison of curriculum and learning.* San Francisco: Jossey-Bass.

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

Shafer, M. C. (2014). The impact of a standards-based mathematics curriculum on classroom instruction and student performance: the case of mathematics in context. In *Mathematics Curriculum in School Education* (pp. 493-514). Netherlands: Springer.

Stein, M. K., Remillard, J., & Smith, M. S. (2007). How curriculum influences student learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319-369). Charlotte, NC: Information Age Publishers.

Strutchens, M. E., Lubienski, S. T., McGraw, R., & Westbrook, S. K. (2004). NAEP findings regarding race and ethnicity: Students' performance, school experiences, attitudes and beliefs, and family influences. In P. Kloosterman & F. K. Lester (Eds.), *Results and interpretations of the 1990-2000 mathematics assessments of the National Assessment of Educational Progress* (pp.269-304). Reston, VA: NCTM

Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education*, *39*(3), 247-280.

Vierra, A., Pollock, J., & Golez, F. (1998). *Reading educational research* (3rd ed.). Upper Saddle River, NJ: Merrill.

Wright, D. B. (2005). The art of statistics: A survey of modern statistics. In P. A. Alexander & Winne (Eds.), *Handbook of educational psychology* (2nd ed.). Mahwah, NJ: Erlbaum.

◈◇◈