Educational Research
Institute of America
www.eriaonline.com • email: info@eriaonline.com
An affiliate of Beck Evaluation & Testing Associates, Inc.

A STUDY OF THE INSTRUCTIONAL EFFECTIVENESS OF *MATH EXPRESSIONS*

Report Number 402
June 2011

Table of Contents

ABSTRACT

*To assess the instructional effectiveness of the Houghton Mifflin Harcourt **Math Expressions** at the Intermediate grade levels, researchers from Educational Research Institute of America (ERIA) assessed student math achievement in grades 3 and 5 for students who used the program over approximately a four month period in the spring semester of the 2010-2011 academic year.*

The assessments for this study were developed by researchers at ERIA. The study and the assessments covered 4 units of study from the grade 3 and grade 5 programs. The reliability and validity evidence for the tests are reported in the study and indicate that the posttest had sound psychometric properties.

All participating teachers either volunteered to participate in the study or were asked to participate by school administrators. An examination of the demographic characteristics of the 3 participating schools indicates the schools are typical of average schools in small towns and urban fringe communities.

*The results showed that the **Math Expressions** classes made significant gains over the course of the tryout period. The effect size was large. The results also show consistent evidence that the **Math Expressions** program is equally effective with students of high and low achievement levels.*

Overview of the Study

International studies have shown that students in a significant number of countries outperform American students on assessments of mathematics (e.g., Provasnik, Gonzales, & Miller, 2009). Because of this international achievement gap, researchers have focused on creating math programs that are highly effective. For twenty-five years, Dr. Karen Fuson, Professor Emeritus of Education and Psychology at Northwestern University, researched effective methods of teaching and learning mathematics which eventually lead to a partnership with Houghton Mifflin Harcourt to development a mathematics program based on Dr. Fuson's research.

In the fall of 2005, Math Expressions was welcomed into classrooms across the nation. Teachers, students, and supervisors quickly saw that by combining aspects of traditional approaches with the most powerful elements of reform teaching. Over this time, experimental research has show that Math Expressions is a highly effective program for improving student math achievement at the primary grade levels (Agondi, Harris, Thomas, Murphy, & Gallagher, 2010), but little research has investigated student math achievement among elementary school students at the intermediate grade levels (i.e., grades 3-5).

Therefore, Educational Research Institute of America (ERIA) was commissioned by Houghton Mifflin Harcourt to conduct an exploratory study of the Math Expressions program with grade 3 and grade 5 students to determine the effectiveness of the program at those two grade levels. This report describes the results of this semester long efficacy study conducted with grade 3 and grade 5 students to determine the impact of Math Expressions ©2011, a Kindergarten through grade 6 mathematics program.

Research Questions

The following research questions guided the design of the study and the data analyses:

1. Is ***Math Expressions*** effective in improving the mathematics skills and problem solving strategies of elementary and middle grade level students?
2. Is ***Math Expressions*** **equally** effective in improving the mathematics skills and problem solving strategies of lower and higher pretest scoring students

Design of the Study

The program's efficacy was evaluated using a pretest/posttest design. Before program instruction, students were administered a pretest designed to cover four instructional units which constituted the content of the ***Math Expressions*** program used for this study. The units included were chosen to match to the extent possible the standards established by the National Council of Teachers of Mathematics (NCTM).

Three grade 3 teachers from 1 school taught units 4, 5, 6, and 7 from the grade 3 **Math Expressions** program. Five grade 5 teachers from 2 different middle schools also taught units 4, 5, 6, and 7 from the **Math Expressions** program. The grade 3 program consisted of a total of 14 units and the grade 5 consisted of a total of 12 units. The instructional program did not get underway until the end of January and instruction was completed by the beginning of May. Thus, it was not possible to cover half of the program. About a third or more of the program was completed at each grade level. Pretests were administered slightly after mid-January and posttests were administered during the first half of May.

Instructional Approach under Study

Following is a description of the program provided by the publisher:

> *Math Expressions is a complete Kindergarten–Grade 6 mathematics curriculum that offers new ways to teach and learn mathematics. Combining the most powerful elements of standards-based instruction with the best of traditional approaches, Math Expressions uses objects, drawings, conceptual language, and real-world situations to help students build mathematical ideas that make sense to them.*
>
> *Math Expressions balances deep understanding with essential skills and problem solving. Students invent, question, and explore, but also learn and practice important math strategies. Through daily Math Talk, students explain their methods and, in turn, become more fluent in them. Every Math Expressions lesson includes intervention, on level, and challenge differentiation to support classroom needs. In addition, leveled math writing prompts provide opportunities for in-depth thinking and analysis, and help prepare students for high-stakes tests.*
>
> *Support for English language learners is integrated throughout the program.*

Description of the Research Sample

Table 1 provides the demographic summary of the three schools included in the study. The demographic data for the schools indicate that for the 2009-2010 school year an average of 17% percent of the students were enrolled in free or reduced lunch programs and an average of 4% of the students were classified as minority students. Fourteen percent of the students were classified as students with special education needs. It is important to note that the school data does not provide a description of the make-up of each of the classes that participated in the study. However, the school data does provide general descriptions of the school.

Table 1
Demographic Characteristics
Of the Schools Included in the Study

Location	*Grades*	*Students Enrolled*	*% Students Free/Reduced Lunch Programs*	*% Minority*	*% Students with Special Education Needs*
Small Town	5-8	601	19%	5%	13%
Urban Fringe Mid-Size City	5-6	541	13%	4%	15%
Urban Fringe Mid-Size City	1-4	391	19%	2%	13%
AVERAGE		512	17%	4%	14%

Description of the Assessments

The pretest and posttest used in the study were developed by mathematics curriculum experts employed by ERIA. The assessments were created to match the math instruction included in the four units at each grade level which were the content focus for the study. The same test was used as the pretest and the posttest. However, the order of items was scrambled for the posttest. In addition, neither students nor teachers were aware that the same items would be used as both the pretests and posttests. Research has shown that the effect of the pretest is minimal when at least two weeks elapse between pretesting and posttesting and that examinees are not aware that the same items will appear on the posttest (Karlin & Jolly, 1970). Using the same items also assures that the two tests will be of equal difficulty since it is impossible to develop two independent forms of a test that are of equal difficulty without conducting extensive test tryouts and item analysis studies to equate item and test difficulties.

The tests covered the four instructional units taught at each grade level and were developed to respond to the following emphases:

- Innovative items that call for actual performance on the part of students that encourage divergent thinking and problem solving
- Emphasis on thinking skills
- Alignment with the NCTM Standards and the State Common Core Standards

The test included 40 multiple-choice items. Table 2 provides the basic test statistics for the pretest and posttest. Table 2 shows that the reliability of the pretest at both grades 3 and 5 were lower than the posttests indicating the probability that more students were guessing answers prior to instruction. The reliabilities for both the pretests and posttests at both grades 3 and 5 were sufficiently high for data analyses. The highest individual student score on the pretest at grade 3 was 34 out of 40 items correct and the highest individual student score on the posttest was 39 out of 40 items correct. At grade 5 the highest score out of a total of 40 test items was 36 and on the posttests the highest student score was 40.

Table 2
Pretest/Posttest Statistics for the *Math Expressions* Assessments for Grades 3 and 5

Test	*Number of Items*	*Mean Score*	*Standard Deviation*	*KR 20*	*SEM*
Grade 3 Assessment					
Pretest	40	20	6.4	.81	
Posttest	40	29	6.2	.84	
Grade 5 Assessment					
Pretest	40	16	5.0	.72	
Posttest	40	27	6.6	.85	

Data Analyses

Data analyses and descriptive statistics were computed for the Mathematics tests developed for each of the total group of 3rd and 5th grade students. Raw scores were converted to standard scores using a mean of 300 and a standard deviation of 50. This was done so the scores approximated a more normal distribution.

The $\leq$.05 level of significance was used as the level at which increases would be considered statistically significant for all of the statistical tests

The following statistical analyses were conducted to compare students' pretest standard scores to posttest standard scores:

- A paired comparison *t*-test was used to compare the pretest mean standard scores with the posttest mean standard scores for all students.
- A mixed model analysis of variance was computed to determine if there was a significant interaction between pretest/posttest scores for the following subgroups:
    - Lower and higher pretest scores

An effect-size analysis was computed for paired *t*-test comparisons. Cohen's *d* statistic was used to determine the effect size. This statistic provides an indication of the strength of the effect of the treatment regardless of the statistical significance. Cohen's *d* statistic is interpreted as follows:

.2 = small effect
.5 = medium effect
.8 = large effect

All Grade 3 Students

Researchers at ERIA conducted a paired comparison *t*-test to determine if the difference from pretest standard scores to posttest standard scores was statistically significant. For this analysis, researchers were able to match the pretest and posttest scores for 66 students. Students who did not take both the pretest and the posttest were not included in the analysis.

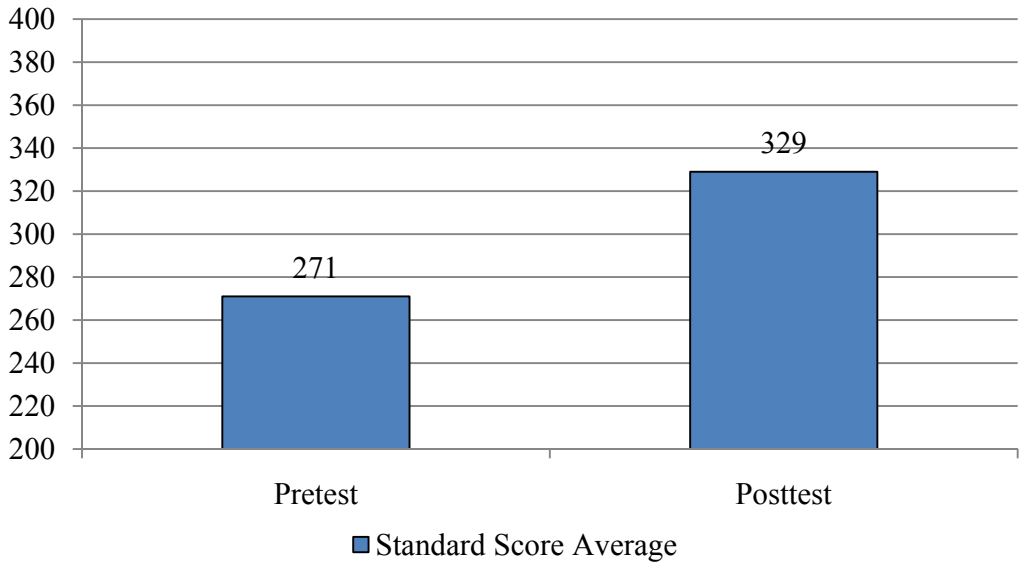
Table 3 shows that the average standard score on the pretest was 271 and the average standard score on the posttest was 329. The increase was statistically significant (<.0001), and the effect size was large.

Table 3
Grade 3 Total Group Paired Comparison *t*-test Results
Pretest/Posttest Comparison of Standard Scores

Test	*Number Students*	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Standard Score (Pretest)	66	271	40.8	16.383	≤.0001	1.40
Standard Score (Posttest)	66	329	41.1			

Figure 1 shows the standard score increase for all grade 3 from pretesting to posttesting was 58 standard score points.

Figure 1
Grade 3 *Math Expressions* Students
Standard Score Changes from Pretesting to Posttesting



Pretest Group Comparison for Grade 3 Students

In order to test for the significance of pretest score difference, a Mixed Model Analysis of Variance (ANOVA) was conducted with higher pretest/lower pretest group as the between subject variable and pretest and posttest scores as the within subject variable. Table 4 shows that the within subject variable (pretest/posttest scores) was statistically significant (≤.0001). The interaction of higher pretest/lower pretest group and the pretest/posttest was not statistically significant.

Table 4
Results of Mixed-Model ANOVA with Higher/Lower Pretest Group a Between-Subject Variable and Test (Pretest/Posttest) a Within-Subject Variable

Test	***Mean Square***	**F-*test***	***Significance***
Pretest/Posttest Effects	109599	276.783	≤.0001
Pretest Group X Pretest/Posttest Interaction	1200	3.030	Non-Significant

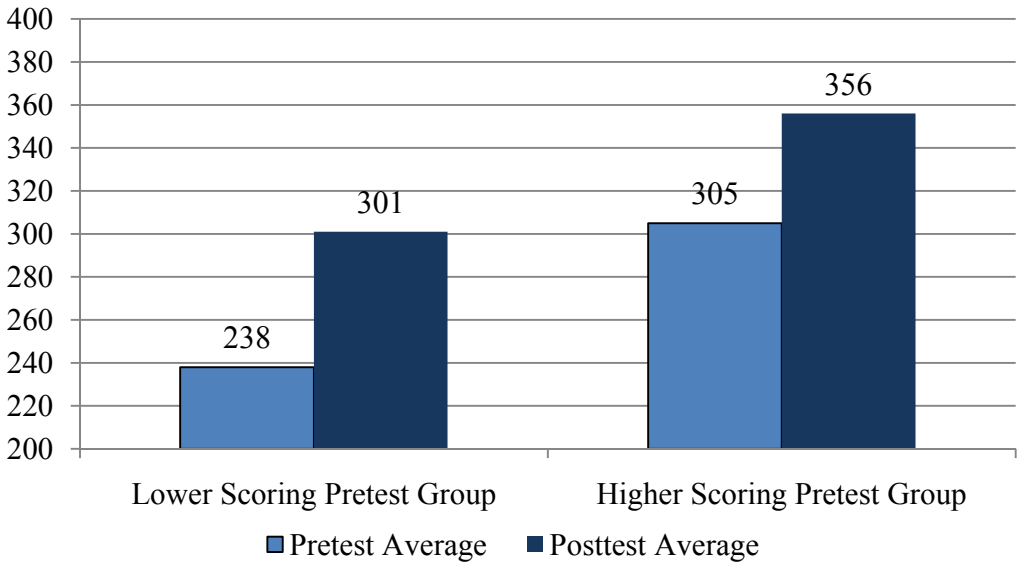
Despite the non-significant effect of the pretest grouping, dependent sample t-tests were computed to determine if significant differences existed between pretest groups from pretesting to posttesting. Table 5 shows that the dependent sample *t*-tests for the pretests and posttests for the high and low pretest groups were both statistically significant (≤.0001). The effect sizes for the both lower and higher pretest scoring groups were large.

Table 5
Paired Comparison t-Tests for Grade 3 Students
Pretest/Posttest Standard Scores Analyzed by Pretest Group

Standard Score	*Number Students*	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Lower Pretest Group						
Pretest	33	238	18.7	12.071	≤.0001	2.24
Posttest	33	301	36.3			
Higher Pretest Group						
Pretest	33	305	26.4	11.486	≤.0001	2.11
Posttest	33	356	23.4			

Figure 2 shows the standard score increase for the lower and higher pretest scoring groups from pretesting to posttesting. The increase was 63 standard score points for the lower pretest scoring students and 51 standard score points for the higher pretest scoring students.

Figure 2
Grade 3 *Math Expressions* Students
Standard Score Changes from Pretesting to Posttesting for Lower and Higher Pretest Scoring Students



All Grade 5 Students

Researchers at ERIA conducted a paired comparison *t*-test to determine if the difference from pretest standard scores to posttest standard scores was statistically significant. For this analysis, researchers were able to match the pretest and posttest scores for 190 students. Students who did not take both the pretest and the posttest were not included in the analysis.

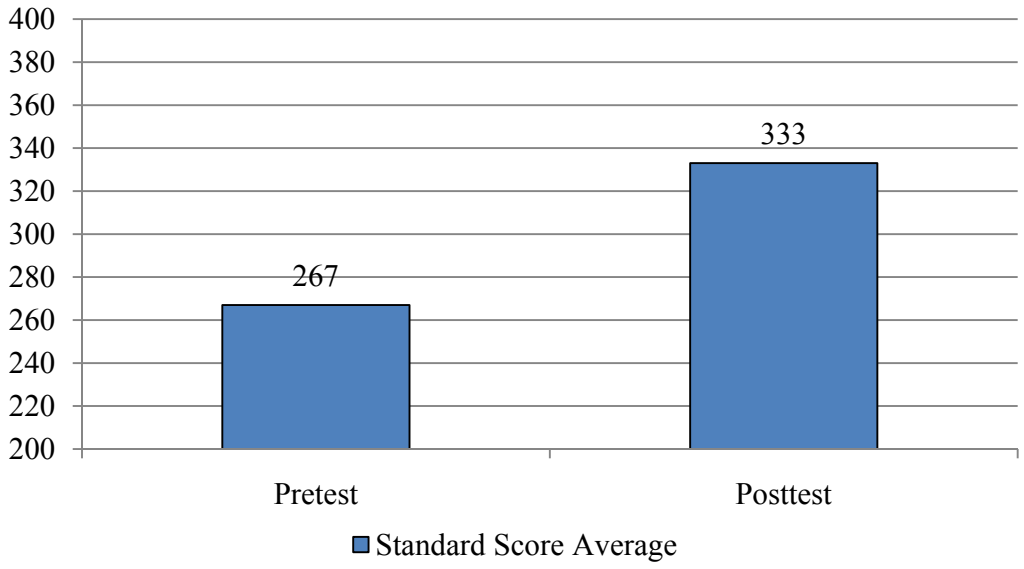
Table 6 shows that the average standard score on the pretest was 267 and the average standard score on the posttest was 333. The increase was statistically significant (<.0001), and the effect size was large.

Table 6
Grade 5 Total Group Paired Comparison *t*-test Results
Pretest/Posttest Comparison of Standard Scores

Test	*Number Students*	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Standard Score (Pretest)	190	267	32.0	23.936	≤.0001	1.78
Standard Score (Posttest)	190	333	42.2			

Figure 3 shows the standard score increase for all grade 5 from pretesting to posttesting was 66 standard score points.

Figure 3
Grade 5 *Math Expressions* Students
Standard Score Changes from Pretesting to Posttesting



Pretest Group Comparison for Grade 5 Students

In order to test for the significance of pretest score difference, a Mixed Model Analysis of Variance (ANOVA) was conducted with higher pretest/lower pretest group as the between subject variable and pretest and posttest scores as the within subject variable. Table 7 shows that the within subject variable (pretest/posttest scores) was statistically significant (≤.0001). The interaction of higher pretest/lower pretest group and the pretest/posttest was also statistically significant (≤.002).

Table 7
Results of Mixed-Model ANOVA with Higher/Lower Pretest Group a Between-Subject Variable and Test (Pretest/Posttest) a Within-Subject Variable

Test	***Mean Square***	**F-*test***	***Significance***
Pretest/Posttest Effects	418276	600.908	≤.0001
Pretest Group X Pretest/Posttest Interaction	7121	10.230	≤.002

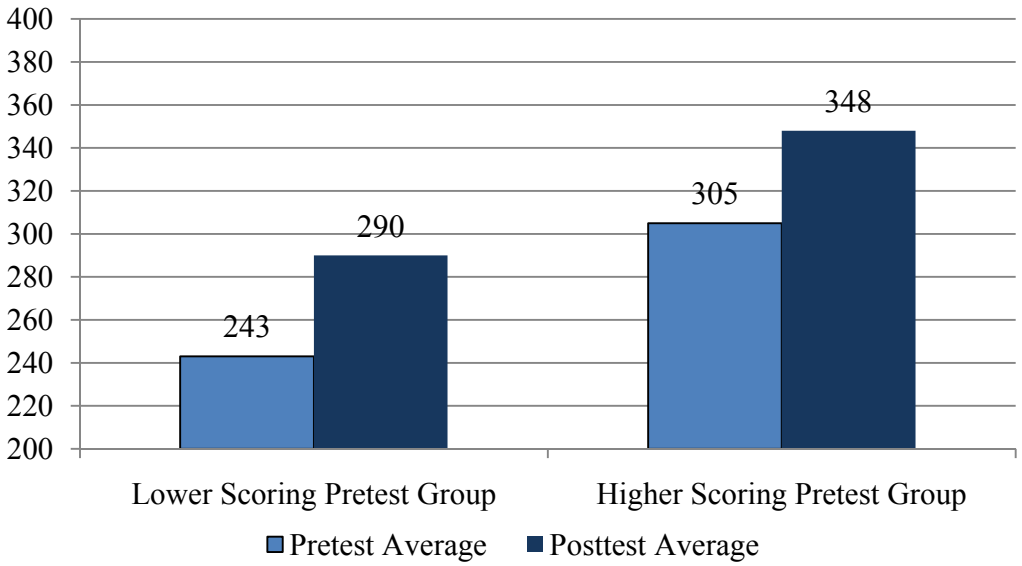
Based on the significant effect of the pretest grouping, dependent sample t-tests were computed to determine if significant differences existed between pretest groups from pretesting to posttesting. Table 8 shows that the dependent sample *t*-tests for the pretests and posttests for the high and low pretest groups were both statistically significant (≤.0001). The effect sizes for the both lower and higher pretest scoring groups were large.

Table 8
Paired Comparison t-Tests for Grade 5 Students
Pretest/Posttest Standard Scores Analyzed by Pretest Group

Standard Score	*Number Students*	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Lower Pretest Group						
Pretest	95	243	12.2	19.956	≤.0001	2.80
Posttest	95	318	35.9			
Higher Pretest Group						
Pretest	95	290	28.3	14.809	≤.0001	1.60
Posttest	95	348	43.0			

Figure 4 shows the standard score increase for the lower and higher pretest scoring groups from pretesting to posttesting. The increase was 47 standard score points for the lower pretest scoring students and 43 standard score points for the higher pretest scoring students.

Figure 4
Grade 5 *Math Expressions* Students
Standard Score Changes from Pretesting to Posttesting for Lower and Higher Pretest Scoring Students



Conclusions

This study sought to determine the effectiveness of ***Math Expressions*** with grade 3 and grade 5 students by comparing the performance of students using the program for about four months using four units of instruction at each grade level.

The following research questions guided the study.

1. Is ***Math Expressions*** effective in improving the mathematics skills and problem solving strategies of grade 3 and grade 5 students?

2. Is ***Math Expressions*** **equally** effective in improving the mathematics skills and problem solving strategies of:
    a. Lower and higher pretest scoring students

Question 1: Is ***Math Expressions*** effective in improving the mathematics skills and problem solving strategies of elementary and middle grade level students?

A valid and reliable mathematics test was used as the pretest and posttest instrument. Paired comparison statistical tests showed that the grade 3 and grade 5 students who were included in the ***Math Expressions*** classes increased their scores statistically significantly and the effect sizes were large.

Question 2: Is ***Math Expressions*** equally effective in improving the mathematics skills and problem solving strategies of students whose achievement was lower or higher on the pretests.

There was a difference favoring lower pretest scoring students to higher pretest scoring students. That is, low pretest scoring differences made greater gains than the higher pretest scoring students. However, both groups made statistically significant gains and the effect sizes for both the lower pretest scoring group and the higher pretest scoring groups were large. In sum, the program was effective with both high achieving and low achieving groups.

In conclusion, the results of this study of four units of study indicates that the Math Expressions program leads to improved mathematics achievement over the course of a four month tryout period. Further, analyses of sub-group performance of students using Math Expressions show that the program does not significantly favor one group over another. In brief this study provides strong evidence of the instructional effectiveness of the Math Expressions program with elementary and middle school grade students.

References

Agondi, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement Effects of Four Early Elementary School Math Curricula: Findings for First and Second Graders (NCEE 2011-4001*). Washington, DC: U.S. Department of Education

Provasnik, S., Gonzales, P., and Miller, D. (2009). *U.S. Performance Across International Assessments of Student Achievement: Special Supplement to The Condition of Education 2009 (NCES 2009-083*). Washington, DC:, U.S. Department of Education.

Karlin, R. & Jolly, H. (1970) The Use of Alternate Forms of Standardized Reading Tests. In R. C. Farr (Ed.), *Measurement and Evaluation of Reading* (pp.192-205). New York, NY: Harcourt Brace & World.