Journal for Research in Mathematics Education
2000, Vol. 31, No. 3, 277–295

Achievement Results for Second and Third Graders Using the *Standards*-Based Curriculum *Everyday Mathematics*

Karen C. Fuson, William M. Carroll, and Jane V. Drueck
Northwestern University

Students using *Everyday Mathematics* (EM), developed to incorporate ideas from the NCTM *Standards*, were at normative U.S. levels on multidigit addition and subtraction symbolic computation on traditional, reform-based, and EM-specific test items. Heterogeneous EM 2nd graders scored higher than middle- to upper-middle-class U.S. traditional students on 2 number sense items, matched them on others, and were equivalent to a middle-class Japanese group. On a computation test, the EM 2nd graders outperformed the U.S. traditional students on 3 items involving 3-digit numbers and were outperformed on the 6 most difficult test items by the Japanese children. EM 3rd graders outscored traditional U.S. students on place value and numeration, reasoning, geometry, data, and number-story items.

Key Words: Achievement; Arithmetic; Curriculum; Early childhood, K–4; Longitudinal studies; Problem solving; Reform in mathematics education

The mathematics education community, stimulated by new economic and technological contexts and by research on students' mathematical thinking, has called for substantial changes in the nature of elementary school mathematics classroom instruction (National Council of Teachers of Mathematics [NCTM], 1989, 1991, 1995). In contrast to traditional textbook instruction focused primarily on rote learning and practice of skills, instruction is envisioned through which students construct meaning for the mathematical concepts and procedures they are investigating and engage in meaningful problem-solving activities (e.g., Cobb & Bauersfeld, 1995; Hiebert et al., 1996; Lampert, 1991). This student construction of mathematical knowledge is facilitated by teachers who elicit, support, and extend children's mathematical thinking (Fraivillig, Murphy, & Fuson, 1999); promote discussions (e.g., Schifter & O'Brien, 1997); use meaningful representations of mathematical concepts (Fuson, Smith, & Lo Cicero, 1997; Fuson, Wearne, et al., 1997); and encourage use of alternative solution methods (Carpenter & Fennema, 1991; Hiebert & Carpenter, 1992). However, results from the recent Third International Mathematics and Science Study (TIMSS) indicate that the U.S. curriculum continues to be an "underachieving curriculum" compared to the math-

---

The research reported in this article was supported by the National Science Foundation (NSF) under Grant ESI 9252984. The opinions expressed in this article are those of the authors and do not necessarily reflect the view of NSF.

ematics curricula in higher achieving nations and that instruction in the United States is still more likely to focus on practice of skills than on understanding (McKnight et al., 1989; Peak, 1996; Stigler, 1997).

A number of U.S. researchers investigating the progress of students experiencing meaning-based instruction have reported positive effects on students' understanding and achievement (Carpenter, Franke, Jacobs, Fennema, & Empson, 1998; Cobb, Wood, Yackel, & Perlwitz, 1992; Fuson, Smith, & Lo Cicero, 1997; Fuson, Wearne, et al., 1997). For example, when compared with students in traditional textbook-based classes, students in Cobb et al.'s Problem-Centered Mathematics Project scored significantly higher on measures of conceptual understanding as well as on standardized tests (Wood & Sellers, 1997). These students also saw mathematics as a more purposeful and understandable activity than did students using traditional approaches. Carpenter, Fennema, and colleagues have reported similar gains for Cognitively Guided Instruction in problem solving and conceptual understanding (Carpenter et al., 1998). Others have reported strong gains in students' conceptual understanding and use of calculation methods when students are actively involved in activities that make mathematics meaningful (Fuson, Smith, & Lo Cicero, 1997; Fuson, Wearne, et al., 1997; Hiebert & Wearne, 1993).

With support from the National Science Foundation and other sources, a number of mathematics educators have developed elementary mathematics programs to attempt to incorporate this research on learning and teaching into a full-scale curriculum. Although these new curricula differ in design and in details, they all were developed to incorporate the ideas of the NCTM *Standards* (1989, 1991, 1995). One of these curricula in wide use around the country is the University of Chicago School Mathematics Project's elementary curriculum *Everyday Mathematics* (EM). The design of this curriculum generally reflects constructivist theories of learning (Steffe & Cobb, 1988; Steffe & Gale, 1995). Students, frequently working in small groups or pairs, actively explore mathematical ideas. Lessons are designed so that students build upon their substantial informal knowledge by making connections to everyday experiences. To scaffold students' thinking during problem solving and discussions, teachers are advised to use manipulatives such as pattern blocks or the hundreds grid for many lessons. By frequently generating and solving story problems, students build conceptual understanding of number and operations. With respect to computational proficiency, both paper-and-pencil and mental activities are designed to allow students to develop conceptual understandings of the operations, and the standard multidigit algorithms are omitted from the curriculum (in accordance with Kamii, 1989). Students are encouraged to invent and discuss their own solution methods. Research from mathematics education and cognitive science regarding the development of conceptual structures and solution methods has also guided the sequence of topics in the curriculum (e.g., see the literature reviewed in Fuson, 1992).

Along with an emphasis on active learning and conceptual understanding, a guiding principle in the development of EM is that developers of the traditional elementary curricula have seriously underestimated the capabilities of children

(Bell, 1974; Bell & Bell, 1988). The EM curriculum was based on the belief that children can learn far more mathematics, with deeper understanding, than has been expected in more traditional programs. Along with whole number concepts and operations, topics that are usually delayed until the upper elementary grades—such as uses of negative numbers, functions, fractions, mental computation, and geometry—are explored beginning in kindergarten. Calculators, rulers, and other mathematical tools are used throughout the curriculum. Because of the breadth of the mathematics covered, developers have taken a spiral approach through which ideas are continuously reviewed and are practiced frequently in different contexts and with increasing complexity. For example, kindergartners and first graders investigate the properties of polygons using geoboards or shapes constructed with plastic straws, and fourth graders make compass and straight-edge constructions and investigate relationships among geometric properties. Games are frequently used to review and practice skills as well as to introduce new concepts.

Although the EM curriculum was extensively field tested and information from classroom observation, teacher feedback, and student tests was incorporated into the revisions (Hedges & Stodolsky, 1987), no study had followed students for multiple years. In conjunction with their funding of development of the EM 4–6 curriculum, the National Science Foundation funded such a longitudinal study of students in the EM curriculum by an outside investigator familiar with the curricular approach (the first author of this study). During the 1994–1995 school year, first graders ($n = 496$) in six school districts using the EM curriculum were tested and interviewed (Drueck, Fuson, & Carroll, 1999). On a broad range of questions, the performance of EM students exceeded that of U.S. students receiving traditional instruction and matched or exceeded performance of one or both of the East Asian (Taiwanese and Japanese) samples on many of the questions (comparison samples were from Stigler, Lee, & Stevenson, 1990).

In the two studies reported here, these same students are followed in second and third grades. Because whole districts often opt for the adoption of a new curriculum, it was difficult to match EM schools to comparable schools for a 5-year longitudinal study. Therefore, existing studies in relevant areas of mathematics were chosen to provide comparisons. For example, during the first year of the study, items from Stigler et al.'s cross-national study (1990) were used (Drueck et al., 1999). A similar design was used in the two studies here (i.e., items were selected because they are considered important in new mathematics curricula, they were taken from tests like the National Assessment of Educational Progress [NAEP] that reflect some consensus about the type of mathematics that students should know, or they were chosen from cross-national comparisons).

In Study 1 we followed the progress of EM second graders on developing concepts related to whole numbers and to multidigit computation. For comparison, assessment items were drawn from a study of U.S. and Japanese second graders (Okamoto, Miura, & Tajika, 1995; Okamoto, Miura, Tajika, & Takeuchi, 1995). In Okamoto et al.'s study, two subtests were constructed, one to assess number sense and the other to assess mathematics achievement, chiefly in computation. Given the "invented algorithm" approach taken by EM, student achievement in each of these areas was of interest. The cross-national nature of the Okamoto et al. study also provided a follow-up to the cross-national aspect of our first-grade study.

In Study 2 we followed the progress of EM third graders in their understandings and uses of whole number concepts and computation together with other mathematical topics, such as geometry and measurement. Results from the fourth NAEP as well as test data from Wood and Cobb (1989) provided a basis for comparing EM students to other U.S. students in several mathematical areas. Although the EM longitudinal sample was not a random sample as was the NAEP sample, the EM sample was selected to represent students and schools from a wide range of backgrounds. Items from the Wood and Cobb Cognitively Based Elementary School Mathematics Test were selected because this test was devised to assess conceptual understanding. It was developed as part of a meaning-focused research project, the Problem-Centered Mathematics Project, and thus reflects mathematical performance valued in those project classes. Assessments in both Study 1 and Study 2 included additional items, including some performance-based questions, that represent other aspects of the EM curriculum.

Formal and informal interactions with the Grade 2 and Grade 3 teachers involved in the study, their principals, and school or district mathematics coordinators indicated that all teachers in each grade used the EM curriculum as their only curriculum—with the exception of one teacher who also used material from a textbook. Data of students from this teacher were included because many teachers supplement any given curriculum, and these data would be biased against the main direction of the results. In these and all interactions, we made clear that we were outside researchers examining strengths and weaknesses of EM and were not representatives for the curriculum itself. Determining how teachers were using the curriculum is a complex issue and would have required more substantial classroom observation and teacher interviews than were allowed for in these studies. We were able to make only one videotaped classroom observation and to hold one teacher post-observation interview for each teacher. Observed lessons were selected to be ones of central importance for the grade level and to permit the display of EM practices (such as discussing children's solution methods) in the lesson. Teachers were randomly assigned to be observed while they taught the selected five lessons on word-problem solving and multidigit addition and subtraction in Grade 2 and the six lessons on addition, subtraction, multiplication, division, and decimals in Grade 3.

The whole-class portions of the classroom observations were coded on a scale of meaning-based classroom practices constructed in consultation with prominent researchers. The most striking strengths identified were the degree to which children were engaged in the learning process and the extent to which teachers established a safe environment in which students could explore and discuss their mathematical thinking (Mills, 1996; Mills, Wolfe, & Brown, 1997). Almost all teachers established classrooms that appeared to have safe and supportive climates. Most children were actively engaged in the learning process and appeared to enjoy learning mathematics. In most classrooms, at least a few children made contributions to the class on their own initiatives. These aspects all relate to recent results concerning aspects of classroom practices that support learning. Stipek et al. (1998) reported that children who enjoyed mathematics learned more than those who did not and that students of teachers who supported learning and effort and encouraged autonomy showed more gains in conceptual understanding than children whose teachers did not engage in these practices.

Additional informal evidence validating the above characteristics in EM classrooms came from extensive conversations with several Grade 3 teachers. They reported conversations among Grade 3 teachers in their buildings concerning differences they all noticed in children who had used EM in Grades 1 and 2. They found that children entered their classes liking mathematics more than in previous years (e.g., "The EM children really look forward to math class"). The children also expected teachers to ask them how they solved a problem, not just to report their answers ("If we don't ask how children solved a problem, they'll just volunteer their method").

STUDY 1

Method

Participants

At the end of the school year, 392 second graders in 22 classes were tested. Of these students, 343 students had been in the original first-grade longitudinal sample. Because we were evaluating the longitudinal effect of the EM curriculum, only the scores of these original 343 students are discussed in this analysis. The 11 schools included urban, suburban, and rural or small-town schools, and the student populations ranged from low-income to affluent. Two classes were Spanish-speaking bilingual classes.

Test, Items, and Procedure

Whole-class tests were administered by a researcher from the Northwestern University Longitudinal Study in April or May of second grade. Each question on the 45-item test was read aloud while students followed along in their test books. Questions were read twice and repeated as necessary, and students were allowed sufficient time to complete each item. Test administration took approximately 60 minutes.

For comparative purposes, a subset of the questions was taken from the Okamoto et al. study (Okamoto, Miura, & Tajika, 1995; Okamoto, Miura, Tajika, & Takeuchi, 1995); 10 items were taken from their number-sense subtest and 14 from their mathematics-achievement subtest. These questions were presented in the same order as in the original study as part of our class test. Okamoto et al.'s study included 29 U.S. second graders attending a middle- to upper-middle-class school in the San

Francisco area and 33 Japanese second-grade students attending a middle-class public school in Tokyo. Because the questions were drawn from the texts at both schools, the test was considered to be "curriculum fair." Although these questions covered only a portion of the mathematics in the EM curriculum (e.g., no geometry or data items were included), all were on topics that were covered in the second-grade EM curriculum. One caveat is that although symbolic computation was tested, this topic had not been given much emphasis in the EM curriculum. Instead, students were more likely to have carried out computations in solving a story problem or as part of a larger activity. Furthermore, Okamoto et al.'s students were middle class to upper-middle class whereas the EM sample was more heterogeneous. Although these differences somewhat complicated direct comparisons, both were biased against EM. Because χ^2 tests were done for the individual items, a more conservative .01 level of significance (instead of .05) was used, $\chi^2(1) \geq 6.64$.

Results and Discussion

Table 1 shows the results on the number-sense test for U.S. and Japanese students in Okamoto et al.'s study and for EM students in this study. EM students scored significantly better than the U.S. traditional students on two items and lower than both the U.S. and Japanese students on one item. No other differences were significant. EM students were outscored on the question "How many numbers are there between 6 and 2?" However, the question is somewhat ambiguous; EM students perhaps interpreted it as "How many *steps* are there between 2 and 6?" or "What is the difference between 6 and 2?" An error analysis showed that 51% of EM

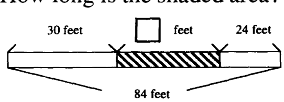
Table 1
Grade 2: Percentages Correct on Number-Sense Test

		Okamoto et al. samples[a]	
Item	EM $n = 343$	U.S. $n = 29$	Japanese $n = 30$
1. Which number is closer to 28: 31 or 22?	88	69*	87
2. How many numbers are between 2 and 6?	35	66*	60*
3. What number comes 4 numbers before 60?	72	66	93
4. What is the smallest 2-digit number?	62	79	47
5. What number comes 10 after 99?	64	59	43
6. What number comes 9 after 999?	41	14*	27
7. Which difference is bigger: between 6 and 2 or between 8 and 5?	46	62	47
8. Which difference is smaller: between 99 and 92 or between 25 and 11?	48	55	40
9. What is the smallest 5-digit number?	43	28	27
10. How much is 301 take away 7?	39	17	33
Mean	54	52	50

[a]These samples are from Okamoto, Miura, & Tajika (1995). The U.S. students were middle class to upper-middle class and used a traditional textbook approach.
*On the chi-square test, significantly different from the EM sample with $p \leq .01$.

students gave the answer *four*, indicating that they had interpreted the question in one of these ways. The only other question answered correctly by fewer than 40% of the EM students was "How much is 301 – 7?" However, this percentage correct was higher than for either the U.S. comparison or Japanese students. Results on the mathematics-achievement test showed a pattern different from the number-sense results, with the Japanese students scoring near ceiling on most items and the EM students scoring between the Japanese and the U.S. comparison students (see Table 2). The Japanese students scored significantly higher than the EM students on the six most advanced items. EM students scored significantly higher than the traditional U.S. students on six problems (four if problems 1a, 1b, and 1c are counted as a single problem); these six items involved knowledge of patterns, addition, and subtraction of tens.

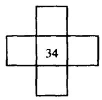
Table 2
Grade 2: Percentages Correct on Mathematics Achievement Test

		Okamoto et al. samples[a]	
Item	EM $n = 343$	U.S. $n = 29$	Japanese $n = 30$
1. Fill in the missing numbers: __, 630, 640, 650, __, __, 680			
1a. 620	95	59*	100
1b. 660	96	55*	100
1c. 670	94	55*	100
2. 67 + 5	87	66*	96
3. 80 – 7	67	76	96*
4. 600 + 100	94	35*	92
5. 110 – 40	50	21*	84*
6. 2 × 3	78	79	100
7. 4 × 1	78	76	100
8. 6 × 4	53	52	92*
9. 1 × 5	77	72	100
10. 296 + 604	54	69	88*
11. 536 – 127	26	41	88*
12. How long is the shaded area? 	24	10	56*
Mean	70	55	92

[a]U.S. and Japanese samples are from Okamoto, Miura, & Tajika, 1995.
*On the chi-square test, significantly different from the EM sample with $p \leq .01$.

Results on other items given to the EM students are shown in Table 3. More than three fourths of the EM second graders correctly wrote 3- and 4-digit numbers, and even when place values were given out of order, as on Problems 5 and 6, two thirds of the students correctly wrote the 3-digit number (the 5-digit number in Item 6 is advanced for second graders but was still answered correctly by 38% of the EM second graders).

Table 3
Grade 2: Percentages Correct on Additional Items From EM Test

Question	EM $n = 343$
Place value	
1. Write the number *five thousand four.*	76
2. Write the number *three hundred twenty-six.*	85
3. Write the number that is 10 more than 57.	85
4. Write the number that is 100 less than 465.	71
5. Write the number that has 6 tens, 3 ones, 5 hundreds.	66
6. Write the number that has 7 thousands, 8 tens, 5 ten thousands, 1 one, 0 hundreds.	38
7. What is the number that is the same as ten tens?	62
8. Complete the number grid: "Here is a piece of the hundreds grid. Fill in the missing numbers on the grid." 	84
Computation	
9. 36 + 47 (vertical format, no context)	65
10. 72 – 26 (vertical format, no context)	38
11. At the water park, the Loop Slide is 65 feet high. The Tower Slide is 28 feet high. How much shorter is the Tower Slide?	30
12. Jim had 63 crayons. He put 10 in each box.	
a. How many boxes did he fill?	56
b. How many crayons were left over?	65
13. There are 264 children at school. How many teams of 10 could you make with these 264 children?	29
Fractions	
14. Circle 1/4 of the dots. •••• ••••	23
15. Draw a circle around one half of the stars. * * * * * * * * * * * *	65

Because Okamoto et al.'s U.S. students were of higher socioeconomic status than those in the heterogeneous EM sample, EM performance on the 2-digit computation items was compared to that of national norms for comparable individual items on a standardized test (Stanford Achievement Test, Psychological Corporation, 1992). The EM students were above national norms for multidigit addition (65% vs. 50%) and at the norm for multidigit subtraction (38%).

EM first graders had demonstrated strong conceptual knowledge of the fraction *one half.* The item requiring circling half of 12 stars in an uneven array was correctly answered by 32% of the EM first graders compared to 11% of U.S. students in traditional instruction (Drueck et al., 1999). By second grade, the proportion of EM students correctly answering this question more than doubled (to 65%), approaching the 71% of U.S. fifth graders in traditional instruction (Stigler et al., 1990) who answered this question correctly.

The strength of EM students seems to be related to the intended curriculum. Compared to time allotted in a traditional curriculum, more time in the EM curriculum was allotted to discussion of students' strategies, such as various counting strategies. These ideas and skills also were reinforced and practiced through counting exercises (e.g., "Write 10 more than 43") and regular activities involving computation and number comparisons on number lines and number grids. EM students also explored fractions in everyday situations from kindergarten onward. In contrast, the EM curriculum had fewer examples of vertical context-free symbolic computations, items on which the EM students did not outperform traditional U.S. samples.

STUDY 2

Comparison computation problems in Study 1 were largely symbolic because the comparison items were originally presented in that way. However, computation in the EM curriculum is usually embedded in a context such as a story problem or a larger problem-solving activity, so the Grade 2 symbolic items did not present a complete picture of the computational abilities of EM students. Study 2 included both symbolic and contextualized computation problems as well as questions in geometry, data, and reasoning. Third-grade items from the fourth NAEP (Brown & Silver, 1989; Kouba, Carpenter, & Swafford, 1989; Lindquist & Kouba, 1989a, 1989b) and from a cognitively based test for Grade 3 (Wood & Cobb, 1989) were used for comparative purposes. Because of the nature of the tests and their construction by experts in the field, they provided items considered to be important both in new and in traditional U.S. mathematics curricula.

Method

Participants

A whole-class test was administered to 620 third graders in 29 classes. These were in the same districts described in Study 1, with additional students and classes due to the mixing of classes and the influx of students new to the schools. Of this group, 236 were part of the original first-grade sample, and their scores are the focus of this study.

Test, Items, and Procedure

Whole-class tests were administered during the month of May by a researcher from the Northwestern University Longitudinal Study staff. Each test consisted of 33 questions and took approximately 50 minutes to administer. The first questions, taken from the Wood and Cobb test (1989), were administered orally, as they had been on the original test; on the remaining items, students worked independently, although, upon request, questions were read to an individual. Four forms of the test were constructed to increase the number of questions without increasing the test time, and two forms were used in each class. All students answered four of the items, and the remaining questions were each answered by about half the EM students.

Testing was planned so that each question was given to students from the whole range of achievement levels and SES backgrounds.

Of the total 64 questions, 22 were taken from the fourth NAEP for the purpose of comparison. Nine were taken from a third-grade cognitively based mathematics test (Wood & Cobb, 1989) given at the same time of the school year. The results for the Wood and Cobb sample are for traditional and problem-centered students combined, as reported by Wood and Cobb. Additional questions were follow-ups to the second-grade tests or were taken from the third-grade EM curriculum. Several performance-based items reflective of the curriculum were included (e.g., drawing or measuring a line segment of a given length).

The 22 NAEP questions were divided into two subtests for analysis: a Number Concepts and Computation subtest and a Geometry, Data, and Reasoning subtest. Each of these subtests contained 11 questions. These questions were presented in the same format as on the NAEP, either multiple choice or open-ended. Chi-square tests were used to compare performance on all NAEP and Wood and Cobb (1989) items. Because of the number of tests, the more conservative .01 level of significance was used (instead of .05), $\chi^2(1) \geq 6.64$. Because between 10% and 15% of the 18,033 students were tested on each NAEP item (Carpenter, 1989), the NAEP sample was assumed to be 1,800 on each question.

Results and Discussion

Number and Computation

As shown in Table 4, the EM third graders scored higher overall than did third graders in the NAEP comparison group on the Number and Computation test (mean 65% vs. 52%). The difference between groups was significant on six of the items, in each case favoring the EM students (by a mean percentage of 25%). EM students outscored the NAEP group on the two questions that involved place-value knowledge (e.g., "What number is 100 more than 498?"), on all three story problems, and on Item 8, which assessed understanding of the connection between addition and subtraction. Thus, EM students did better on problems that were more conceptual or that involved a context.

On eight of the nine items, EM students scored significantly higher than the Wood and Cobb (1989) students (an economically heterogeneous sample comprised of some students receiving a traditional approach and some students receiving the Wood and Cobb meaning-based approach) (see Table 5). The EM students scored about 20% higher on each of the six number stories (addition, subtraction, multiplication, and division story problems), on a numerical problem with an unknown factor (3 * ____ = 27), and on the unknown-added problem in the context of base-ten blocks.

To assess progress of EM students in computation, we repeated three symbolic computation questions from the second-grade test (Study 1) on the third-grade test and gave a comparable 2-digit subtraction problem (54 – 37). As the results in Table

Table 4
EM Grade 3 and NAEP Grade 4: Percentages Correct on NAEP Number and Computation Items

Question	EM Grade 3 n = 107 to 119[a]	NAEP Grade 4 n = 1,800[b]
Place value		
1. What digit is in the thousands place in the number 43,486?	67*	45
2. What number is 100 more than 498?	80*	43
Symbolic computation (Vertical form except Question 8, which was horizontal)		
3. 57 + 35	79	84
4. 49 + 56 + 62 + 88	60	48
5. 54 – 37	72	70
6. 504 – 306	38	45
7. 242 – 178	62	50
8. If 49 + 83 = 132, which of the following is true? (132 – 49 = 83 is the answer)	56*	29
Computation in number stories		
9. Robert spends 94 cents. How much change should he get back from $1.00?	85*	68
10. Chris buys a pencil for 35 cents and a soda for 59 cents. How much change does she get back from $1.00?	59*	29
11. At the store, a package of screws costs 30 cents, a role of tape costs 35 cents, and a box of nails costs 20 cents. What is the cost of a roll of tape and a package of screws?	77*	58
Mean	65	52

[a]From a total of 236, EM samples varied across various subsamples of 107, 117, and 119. Item samples are available from the authors. [b]A total of 18,033 third graders participated in the fourth NAEP. Only 10% to 15% of these students answered each item (Carpenter, 1989). On Chi-square tests, the NAEP subsample was assumed to be 1,800 on each item.
*On the chi-square test, the EM sample was significantly higher than the NAEP sample, $p \leq .01$.

6 indicate, EM students made progress on both multidigit addition and subtraction. Along with better performance on subtraction, the incidence of the common error of "subtracting the smaller digit from the larger in each column" decreased from 31% for the students in Grade 2 (50% of the 62% incorrect) to only 12% for the students in Grade 3 (43% of the 28% incorrect).

Geometry, Data, and Reasoning

The EM students significantly outperformed the NAEP students on the four geometry items, half the data and graphing items, and the reasoning item (see Table 7). Differences were especially large for the following items: finding the perimeter of a rectangle with length and width shown (50% higher), showing a conceptual understanding of area (36% higher), and using reasoning (35% higher). In fact, EM third graders did as well as or outperformed the seventh graders in the NAEP sample on three of the questions: finding the area of a 6-by-5 rectangle with square units

Table 5
Grade 3: Percentages Correct on Items From Wood and Cobb Test

Question	EM n = 107 to 119[a]	Wood & Cobb n = 191
Number stories		
1. Paul planted 46 tulips. His dog dug up some of them. Now there are 27 tulips left. How many tulips did Paul's dog dig up?	68*	49
2. Sue had some crayons. Then her mother gave her 14 more crayons. Now Sue has 33 crayons. How many crayons did Sue have in the beginning?	76*	50
3. Ann and Stacy picked 31 roses altogether. Ann picked 17 roses. How many roses did Stacy pick?	79*	52
4. Mary, Sue, and Ann sold 12 boxes of candy each. How many boxes of candy did they sell in all?	74*	49
5. There were 48 birds in a tree. Then, 14 flew away and 8 more arrived. How many birds are in the tree?	70*	51
6. In school, 24 children play soccer. Each soccer team has 6 players. How many teams are there?	88*	60
Place value and conceptual addition/subtraction		
1. There are 12 cubes hidden in the box. How many cubes are there altogether? (Drawing shows 4 ten-longs, 7 unit-cubes [base-10 blocks], and a box.)	77	67
2. Some cubes are hidden in the box. There are 57 cubes altogether. How many cubes are hidden? (Drawing shows 2 ten-longs, 2 unit-cubes [base-10 blocks], and a box.)	73*	50
Multiplication and division computation		
1. 3 * ___ = 27	80*	59

[a]Because different forms of the test were given, the number of EM students varied from a total sample of 236 across subsamples of 107, 117, and 119. Item samples are available from the authors.
*On the chi-square test, the EM sample was significantly higher than the Wood and Cobb (1989) sample with $p \leq .01$.

Table 6
Second- and Third-Grade EM: Percentages Correct on Longitudinal Symbolic Computation

Question	End of second grade n = 343	End of third grade n = 236
80 – 7	67	82
110 – 40	50	80
296 + 604	54	78
72 – 26	38	—
54 – 37	—	72

Note. Comparable but different 2-digit subtraction problems were used in the two tests.

shown (56% of students in both groups correct), finding the perimeter of a 4-by-7 rectangle with the dimensions given (67% of EM third graders correct vs. 46% of the seventh graders), and the reasoning question (64% of EM third graders correct vs. 45% of the seventh graders). EM students scored a mean of 23% higher on the three significant data and graphing items.

Table 7
EM Grade 3 and NAEP Grade 4:
Percentages Correct on Geometry, Data, and Reasoning Items

Question	EM Grade 3 $n = 107$[a]	NAEP Grade 4 $n = 1,800$[b]
Geometry		
1. What is the area of this rectangle?		
a. 6-by-5 rectangle with square units shown	56*	20
b. With length and width shown (6 by 5)	19*	5
2. What is the perimeter of this rectangle?		
a. What is the distance around a 4-by-7 rectangle?	23*	15
b. With length and width shown (4 by 7)	67*	17
Data and graphing		
3. Using a graph		
a. Reading bar graph	80	67
b. Comparing information from bar graph	54*	29
c. Combining information from bar graph	46	44
4. Using a table		
a. Reading a table	87*	70
b. Comparing information in a table	60*	34
c. Combining information in a table	63	58
Reasoning		
5. Four cars wait in a single line at a traffic light. The red car is first in line. The blue car is next to the red. The green car is between the white car and the blue car. Which car is at the end of the line?	64*	29
Subtest mean	56	35

[a]Because different forms of the test were given, slightly different numbers of students were tested on different items (EM n = 129 or 107). [b]Nationwide, a total of 18,033 third graders participated in the fourth NAEP. Only 10% to 15% of these students answered each question (Carpenter, 1989). A sample of 1,800 was assumed for the chi-square tests.
*On the chi-square test, the EM sample was significantly higher than the NAEP sample, $p \leq .01$.

Although these problems were presented separately from computation, some of the geometry and data problems obviously involved computation (e.g., finding the perimeter of a rectangle or adding data from a table). Thus, the emphasis of the EM program on problem solving, applications, and computation in a context seems to be effective in reducing the consistent complaint about students in traditional curricula—that even when they master algorithms, they can have difficulty using these algorithms in applied situations (Kouba et al., 1989).

The high scores on the area and perimeter questions indicate that the EM emphasis on meaningful concrete exploration of traditionally underrepresented topics like geometry and measure is effective. Perhaps because EM students use tools (e.g., rulers, tape measures, and pattern-block geometry templates) and manipulatives (e.g., using geoboards to construct rectangles and counting the distance around) that involve area and perimeter, these students are less likely to confuse the two concepts.

SUMMARY

Various efforts are underway nationally to improve the mathematics achievement of U.S. students. The approach taken in the University of Chicago School Mathematics Project and similar meaning-based projects and curricula is an attempt to replace the "underachieving curriculum" (McKnight et al., 1989) with a more ambitious and meaningful mathematics program grounded in solving problems in contexts (rather than mostly symbolic problems), using manipulatives and tools to facilitate children's thinking, and fostering children's mathematical thinking by teachers. Whereas traditional U.S. primary programs have focused on practice of facts and of whole number algorithms, the EM curriculum and other reform programs also include a wider range of mathematical topics as envisioned by the NCTM *Standards* (1989).

Results from the two studies here show positive results for this approach. EM students at Grades 2 and 3 were at normative U.S. levels on multidigit addition and subtraction symbolic computation. On a test of number sense, the heterogeneous EM Grade 2 students scored higher than middle-class to upper-middle-class U.S. traditional-textbook students on two items and matched them on the remaining items, and their scores were equivalent to those of middle-class Japanese students. On a computation test, the Grade 2 EM students outperformed the same U.S. students on three items involving 3-digit numbers. They were, however, outperformed on the six most difficult test items by the Japanese children. Compared to other heterogeneous groups of U.S. students using traditional approaches, EM Grade 3 students scored higher on items assessing knowledge of place value and numeration, reasoning, geometry, data, and solving number stories. EM third graders even showed performance equivalent to or stronger than NAEP seventh graders on a few questions in these areas. Given the generally poor performance of U.S. students in geometry and measurement, such as on the recent TIMSS and sixth NAEP (Kenney & Silver, 1997), these results show the improvements in both understanding and achievement that can be attained with a more ambitious elementary curriculum.

Stipek et al. (1998) found that teachers' practices promoted by motivation researchers and mathematics education reformers (focusing on learning and effort and encouraging autonomy) enhanced students' conceptual understanding. They related this finding to experimental motivation studies in which focusing subjects' attention on mastery (as opposed to performance) contributed to "deep" as opposed to "shallow" processing. Performance on some of the tasks on which EM students outperformed other students provides indirect support for the interpretation that EM students were approaching tasks in a deeper, more engaged way. For example, the reasoning task (Table 7, Item 5) on which the Grade 3 EM students outperformed NAEP Grade 7 students (64% to 45%) is simple if students draw a picture, a deeper form of engagement with the problem. Similarly, number stories are more accessible if students try to understand the underlying situation instead of focusing on key words or on the sizes of numbers (shallow strategies frequently used by students using traditional textbooks).

Children's opportunity to learn was an important issue in interpreting the results of this study. However, several other issues relate to opportunity to learn. EM developers recommend 60 minutes of class time a day, and schools in the study reported scheduling that much time for math—exceeding the more common 45-minute mathematics period. However, this greater time for learning was accompanied by two other important changes: the inclusion of more ambitious topics and the support of learning in the new ways discussed in this report. Topics generally underrepresented or delayed in traditional curricula, such as geometry, fractions, and algebra, were explored at all grades in the EM curriculum. To assist students, mathematical ideas were often presented in real-life contexts and in problem-solving activities. Alternative solution methods were to be elicited and discussed. In brief, a greater opportunity to learn, in terms of both total time and the inclusion of more ambitious topics, was accompanied by activities that made the mathematics meaningful to the students.

Some caveats are also important in interpreting these results. First, we do not claim or show that *Everyday Mathematics* is the only or the best of the new curricula approaches. We suggest only that children learning from the EM curriculum can learn more than children learning from teachers using a traditional curriculum. Second, we are not arguing for the inclusion of any particular topic at any grade. We are concluding only that U.S. children can learn more advanced topics not ordinarily covered in traditional textbooks.

Third, we are not saying that the EM teachers were exemplary teachers but only that their classrooms showed certain characteristics described above. In fact, the classroom coding indicated some areas of relative weakness. Almost all whole-class discourse was teacher-to-student instead of student-to-student; the majority of student responses were brief; descriptions and discussion of solution methods were largely superficial; and few teachers extended student thinking (see Fraivillig et al., 1999; Mills, 1996; Mills et al., 1997, for more details).

Fourth, although EM student computation was at normative levels (i.e., it was as good as performance of students using traditional textbooks), this normative level was not as high as the level in East Asian countries and not as high as one would wish (e.g., at Grade 2, only 38% correct on 2-digit subtraction with regrouping). EM did not "fix" this national computation problem as well as it "fixed" learning in other areas. The reasons for this finding are complex and cannot be summarized briefly (see Mills & Fuson, 1998, for a discussion and more data).

Fifth, and related to all the above, is the issue of breadth versus depth in the topics covered in a given year. This issue is one that needs to be addressed in future research and in discussion within the research community. EM developers deliberately chose a spiral approach in which topics were repeated within a year and across years. Many teachers reported difficulty with this approach because they did not know when to seek mastery of a particular topic by all children. Furthermore, in the first three grades, the one or two teachers in whose classrooms we saw in-depth discussion of student thinking articulated their vision of the curriculum as consisting of a progression or range of solution methods through which they helped all children move (what Simon, 1995, called a "learning trajectory"); they did not view the curriculum as being composed just of the content of the EM lessons. These teachers looked for ways to help children move along throughout the year rather than just in the EM lessons focused on these topics, and they felt comfortable stopping on a given day to follow up on student thinking. Other teachers said that they felt considerable pressure to "cover" or "get through" the EM curriculum because there were so many lessons; in fact, no teacher taught all lessons in any year. Thus, there is conflict resulting from at the same time increased breadth of a curriculum that includes more advanced new topics and the depth required in allowing time for children to discuss their thinking. The trade-offs need to be examined in future research.

One alternative that might be explored in such research is the approach taken in the *Children's Math Worlds* project: Concentrate on more ambitious grade-level goals that are connected to the usual grade-level goals instead of including new topics such as fractions. Teachers can help urban children from poverty backgrounds if the teachers are supported by a curriculum that has ambitious computational goals more in line with East Asian curricula (e.g., 2-digit addition with regrouping in Grade 1) and enables teachers to support children through a learning progression of single-digit (Fuson, Perry, & Ron, 1998) and multidigit methods (e.g., Fuson, 1998). In this project, urban children from poverty backgrounds outperform U.S. children from an economic range of backgrounds and look more like East Asian children in their performance and conceptual understanding (Fuson, 1996; Fuson, Smith, & Lo Cicero, 1997).

The reform movement is under attack nationally as promoting "fuzzy mathematics" and as failing to support traditional grade-level calculation performance. Our results from the most widely used elementary reform curriculum (at the time of the research) do not support such critics. On traditional vertical symbolic multidigit addition and subtraction, EM students performed as well as students using traditional approaches. On a wide range of other mathematically and conceptually demanding tasks, EM students outperformed other groups. Thus, this study provides an existence proof that U.S. students can perform considerably better than they ordinarily do when learning from traditional approaches and that teachers can learn to support such learning through use of a carefully developed curriculum.

REFERENCES

Bell, M. S. (1974). What does "Everyman" really need from school mathematics? *Mathematics Teacher, 67,* 196–202.

Bell, M. S., & Bell, J. B. (1988). *Assessing and enhancing the counting and numeration capabilities and basic operation concepts of primary school children.* Unpublished manuscript, University of Chicago.

Brown, C. A., & Silver, E. A. (1989). Data organization and interpretation. In M. M. Lindquist (Ed.), *Results from the fourth mathematics assessment of the National Assessment of Educational Progress* (pp. 28–34). Reston, VA: National Council of Teachers of Mathematics.

Carpenter, T. P. (1989). Introduction. In M. M. Lindquist (Ed.), *Results from the fourth mathematics assessment of the National Assessment of Educational Progress* (pp. 1–9). Reston, VA: National Council of Teachers of Mathematics.

Carpenter, T. P., & Fennema, E. (1991). Research and cognitively guided instruction. In E. Fennema, T. P. Carpenter, & S. J. Lamon (Eds.), *Integrating research on teaching and learning mathematics* (pp. 1–16). Albany: State University of New York Press.

Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E., & Empson, S. B. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education, 29,* 3–20.

Cobb, P., & Bauersfeld, H. (Eds.). (1995). *The emergence of mathematical meaning: Interaction in classroom cultures.* Hillsdale, NJ: Erlbaum.

Cobb, P., Wood, T., Yackel, E., & Perlwitz, M. (1992). A follow-up assessment of a second-grade problem-centered mathematics project. *Educational Studies in Mathematics, 23,* 483–504.

Drueck, J. V., Fuson, K. C., & Carroll, W. M. (1999). *Performance of U.S. first graders in a reform math curriculum compared to Japanese, Chinese, and traditionally taught U.S. students.* Manuscript submitted for publication.

Fraivillig, J. L., Murphy, L. A., & Fuson, K. C. (1999). Advancing children's mathematical thinking in Everyday Mathematics classrooms. *Journal for Research in Mathematics Education, 30,* 148–170.

Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 243–275). New York: Macmillan.

Fuson, K. C. (1996, April). *Latino children's construction of arithmetic understanding in urban classrooms that support thinking.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Fuson, K. C. (1998). Pedagogical, mathematical, and real-world conceptual-support nets: A model for building children's mathematical domain knowledge. *Mathematical Cognition, 4,* 147–186.

Fuson, K. C., Perry, T., & Ron, P. (1996). Developmental levels in culturally-different finger methods: Anglo and Latino children's finger methods of addition. In E. Jakubowski, D. Watkins, & H. Biske (Eds.), *Proceedings of the eighteenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 347–352). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Fuson, K. C., Smith, S. T., & Lo Cicero, A. M. (1997). Supporting Latino first graders' ten-structured thinking in urban classrooms. *Journal for Research in Mathematics Education, 28,* 738–760.

Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., Carpenter, T. P., & Fennema, E. (1997). Children's conceptual structures for multidigit numbers and methods of multidigit addition and subtraction. *Journal for Research in Mathematics Education, 28,* 130–162.

Hedges, L. V., & Stodolsky, S. S. (1987). *A formative evaluation of Kindergarten Everyday Mathematics* (Evaluation report #86/87-KEM-1). Chicago: University of Chicago School Mathematics Project.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher, 25*(4), 12–21.

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal, 30,* 393–425.

Kamii, C. (with Joseph, L. L.). (1989). *Young children continue to reinvent arithmetic—2nd grade: Implications of Piaget's theory.* New York: Teachers College Press.

Kenney, P. A., & Silver, E. A. (1997). *Results from the sixth mathematics assessment of the National Assessment of Educational Progress.* Reston, VA: National Council of Teachers of Mathematics.

Kouba, V. L., Carpenter, T. P., & Swafford, J. O. (1989). Number and operations. In M. M. Lindquist (Ed.), *Results from the fourth mathematics assessment of the National Assessment of Educational Progress* (pp. 64–93). Reston, VA: National Council of Teachers of Mathematics.

Lampert, M. (1991). Connecting mathematical teaching and learning. In E. Fennema, T. P. Carpenter, & S. J. Lamon (Eds.), *Integrating research on teaching and learning mathematics* (pp. 121–152). Albany: State University of New York Press.

Lindquist, M. M., & Kouba, V. L. (1989a). Geometry. In M. M. Lindquist (Ed.), *Results from the fourth mathematics assessment of the National Assessment of Educational Progress* (pp. 44–54). Reston, VA: National Council of Teachers of Mathematics.

Lindquist, M. M., & Kouba, V. L. (1989b). Measurement. In M. M. Lindquist (Ed.), *Results from the fourth mathematics assessment of the National Assessment of Educational Progress* (pp. 35–43). Reston, VA: National Council of Teachers of Mathematics.

McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., & Cooney, T. J. (1989). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective.* Champaign, IL: Stipes.

Mills, V. L. (1996, April). *Observing second-grade classrooms implementing Everyday Mathematics: What do we see of reform goals?* Paper presented at the annual meeting of the American Educational Research Association, New York.

Mills, V. L., & Fuson, K. C. (1998). *Issues surrounding reform-curriculum computational performance.* Manuscript in preparation.

Mills, V. L., Wolfe, R., & Brown, R. (1997, March). *Supporting reform goals through a reform curriculum: Observing second- and third-grade teachers implementing Everyday Mathematics.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

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. (1995). *Assessment standards for school mathematics.* Reston, VA: Author.

Okamoto, Y., Miura, I. T., & Tajika, H. (1995, April). *Children's intuitive understanding of number and formal mathematics learning: A cross-national comparison.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Okamoto, Y., Miura, I. T., Tajika, H., & Takeuchi, Y. (1995, April). *A developmental study of U.S. and Japanese children's representations of number, place value understanding, and mathematics achievement.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Peak, L. (1996). *Pursuing excellence: A study of the U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in an international context.* Washington, DC: National Center for Educational Statistics.

Psychological Corporation. (1992). *Stanford Achievement Test, Primary 2, Form L.* Harcourt Brace Jovanovich.

Schifter, D. E., & O'Brien, D. C. (1997). Interpreting the standards: Translating principles into practice. *Teaching Children Mathematics, 4,* 202–205.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26,* 114–145.

Steffe, L. P., & Cobb, P. (1988). *Construction of arithmetical meanings and strategies.* New York: Springer-Verlag.

Steffe, L. P., & Gale, J. (1995). *Constructivism in education.* Hillsdale, NJ: Erlbaum.

Stigler, J. W. (1997, April). *Classroom mathematics instruction in Germany, Japan, and the United States: An introduction to the TIMSS videotape classroom study.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Stigler, J. W., Lee, S., & Stevenson, H. W. (1990). *Mathematical knowledge of Japanese, Chinese, and American elementary school children.* Reston, VA: National Council of Teachers of Mathematics.

Stipek, D., Salmon, J. M, Givvin, K. B., Kazemi, E., Saxe, G., & MacGyvers, V. L. (1998). The value (and convergence) of practices suggested by motivation research and promoted by mathematics education reformers. *Journal for Research in Mathematics Education, 29,* 465–488.

Wood, T., & Cobb, P. (1989). *The development of a cognitively-based elementary school mathematics test: Final report.* West Lafayette, IN: Purdue University, School Mathematics and Science Center.

Wood, T., & Sellers, P. (1997). Deepening the analysis: Longitudinal assessment of a problem-centered mathematics program. *Journal for Research in Mathematics Education, 28,* 163–186.

Authors

Karen C. Fuson, Professor, School of Education and Social Policy, Northwestern University, 2115 N. Campus Drive, Evanston, IL 60208-2610; fuson@nwu.edu

William M. Carroll, St. Ignatius School, 1076 Roosevelt Road, Chicago, IL 60608; carroll_w @ignatius.org

Jane V. Drueck, 1928 Chestnut Avenue, Wilmette, IL 60091