A Longitudinal Study of Children in the Everyday Mathematics Curriculum

William M. Carroll

Fall 2001

University of Chicago School Mathematics Project

A Longitudinal Study of Children in the Everyday Mathematics Curriculum

William M. Carroll

Fall 2001

University of Chicago School Mathematics Project

The research described in this report was carried out by Northwestern University with funding from the National Science Foundation (Grant ESI-925984). Statements or conclusions herein do not necessarily assume or suggest policies or beliefs of the National Science Foundation.

Table of Contents

Introduction

Development of the University of Chicago School Mathematics Project's (UCSMP) elementary program began in the mid-1980s with the writing, field testing, and publication of *Kindergarten Everyday Mathematics*. A large body of research conducted by Professor Max Bell (Bell & Bell, 1988), the director of the UCSMP elementary component, and other educational researchers laid the foundation for this problem-solving curriculum:

- Young children begin school with knowledge of mathematics that was not acknowledged by current programs.
- Mathematics programs failed to use and build upon children's informal knowledge of mathematics.
- Compared to mathematics instruction in other nations, the U.S. curriculum was dull and repetitive, with much of the year being spent reviewing old skills and focusing largely on number and symbolic computation.
- Understanding was often improved when children were allowed to investigate and use their own solution methods. Manipulatives and familiar contexts provided scaffolding for this deeper understanding.

Current mathematics programs also failed to recognize the usefulness of technology, such as hand-held calculators, and how this would transform mathematics instruction (Bell, 1974). *Kindergarten Everyday Mathematics* was published in 1987, and funding from GTE, Everyday Learning Corporation, and the authors supported the development of *Everyday Mathematics* (EM) for first through third grade during the period from 1987 through 1993.

During the 1980's, international tests and studies indicated that U.S. students lagged behind their peers in other nations in their knowledge of mathematics and the U.S. mathematics program was characterized as an "underachieving curriculum' (McKnight et al. 1987; Stevenson & Stigler, 1992). Results on national mathematics tests also showed poor results on questions that went beyond simple facts and computation (Lindquist, 1989). This and other research caused the National Council of Teachers of Mathematics and educators to call for new approaches to teaching mathematics, with an emphasis on problem-solving, reasoning, and conceptual understanding.

Following the publication of the NCTM *Standards* (1989, 1991), the National Science Foundation provided funding for the development of several curricula, including EM, to implement these ideas. Along with its support for development of *Everyday Mathematics 4–6*, the National Science Foundation funded a longitudinal study of children using the curriculum. The purpose of the study was to document the mathematical development and achievement of children in a *Standards*-based curriculum, as well as to provide feedback to curriculum developers regarding how to best implement the NCTM Standards. Dr. Karen Fuson of Northwestern University was selected to design and carry out this longitudinal study of *Everyday Mathematics* and outside evaluation.

Chapter 1
Overview and Design of the Longitudinal Study (1992–1993)

Planning for the longitudinal study began during the 1992–1993 school year. During this year, researchers from Northwestern University visited and observed classes in which *Everyday Mathematics* (EM) was used. Because of the longitudinal nature of this study, schools were selected only if they had adopted the EM curriculum through third grade and planned on adopting it through sixth grade[1].

During this year, researchers from Northwestern also analyzed the EM curriculum and its goals, how these goals and lessons coincided with the NCTM *Standards* and other calls for reform in mathematics education, and the degree to which instruction reflected these goals. Based on this analysis, class observation instruments were developed to assess the implementation of EM.

Participants

In selecting participants, there were two major concerns. First, schools were selected so that student backgrounds would provide as diverse a sample as possible. Because control classes were not used, a heterogeneous sample allowed for more valid comparisons to previous studies. Second, so as not to confound results with poor EM implementations (such as teachers who heavily supplemented with drill sheets), first-grade teachers whose teaching exemplified the goals of EM were selected[2].

Following several observations by two researchers from Northwestern University, 24 teachers in 11 schools were selected to participate in the first-

[1] An ideal research situation would have existed if some classes in a school used EM and others did not, providing control classes. However, textbook adoptions were school-wide and, in most cases, district-wide.

[2] In the years that followed, children were generally randomly assigned to classes. No choice in teachers or the level of implementation was possible.

grade study. About 500 students were in the sample. Schools were urban (including three Chicago Public Schools), suburban, and rural / small town, and students came from a wide range of social-economic, racial, and ethnic backgrounds. For example, one of the suburbs was largely working class with a substantial Hispanic population, while in a second suburban school district, approximately one-third of the children were from low-income families. Two of the first-grade classes were Spanish bilingual.

Procedure

The initial first-grade sample of children was tested annually as they progressed from first through fifth grade. Additionally, a random sample of 240 first graders (5 boys and 5 girls from each classroom) was selected from this larger group to participate in individual interviews. Of this original sample of approximately 500 students, 171 were tested and 89 were interviewed annually through fifth grade. Thus, about one-third of the original testing sample and three-eighths of the original interview sample were followed for all of the five years.

Although the research design of the longitudinal study varied from year to year, three overarching research questions were pursued throughout the study.

- **The achievement and mathematical understanding of students:** Each year, whole-class tests were administered to all students involved in the study. (Although all students in classes with participating students were tested, only participating students' tests were scored as part of the study.) To provide for comparisons to normative samples, many of the test and interview questions were selected from previous studies.

- **Teacher change and implementation of a *Standards*-based curriculum:** Because of the differences of EM from traditional mathematics instruction, it was expected that teachers would experience some difficulties in implementing the curriculum. Some of these might be transitory, e.g., learning the routines of the programs or relearning geometric terms, while others might be longer term, e.g., managing a classroom in which children use multiple solution methods rather than standard algorithms. Each year, teachers were interviewed and surveyed about their use of EM. During some years, meetings were also held with groups of teachers.
- **The nature of classroom interactions in a *Standards*-based curriculum:** It was expected that EM classes would differ from more traditional mathematics classes in a number of ways. Rather than practicing facts and computation, children would be involved in problem-solving activities requiring reasoning and discussion. Small group and partner work involving manipulatives and a wide variety of mathematical representations would be typical, and a broader range of mathematical topics would be observed. At each grade level, classes were observed and videotaped or audiotaped. Generally two researchers from Northwestern took part in each observation, one taking live notes and the second observer working a video camera. Notes and videotapes were transcribed and combined for various analyses (Fraivillig, Murphy, and Fuson, 1999).

This report focuses on student achievement results, especially test questions that allow comparisons to students not in the EM program, e.g., items from the National Assessment of Educational Progress at relevant grades. Student achievement results from each year are described consecutively. Classroom observations, implementation issues pertinent to *Standards*-based instruction, and teacher comments about the EM program are discussed in the last chapter.

Published reports on the longitudinal study cited in the references provide further detail on the studies. Additional unpublished reports are available from the UCSMP Elementary Component, 5853 S. Kimbark, Chicago, IL 60637.

Chapter 2
First Grade Study (1993–1994)

Participants and Procedure

Whole-class tests were administered to all first graders in the 24 participating classrooms during November and again in the spring (April/May). Most of the children had used *Kindergarten Everyday Mathematics* the previous year. Four hundred ninety-six students took the fall test, and 503 took the spring test. Additionally, 240 of the students were selected to take part in individual interviews. Students in this sub-sample were interviewed annually throughout the longitudinal study. The 11 schools (24 classes) in the sample are described in Table 2.1.

Table 2.1: Description of the 11 schools participating in the longitudinal study

District	Number of classes (schools) participating	Description of district	Percent of low-income students at each school
A	5 (3 schools)	Large urban with majority of students low-income	29%; 35%; 59%
B	6 (2 schools)	Small city bordering on Chicago; very diverse both in ethnicity and SES	18%; 48%
C	3 (2 schools)	Working-class suburb with large Hispanic population (2 Spanish-bilingual classes)	50%; 60%
D	5 (2 schools)	Largely affluent suburb with pockets of low-income families	7%; 13%
E	4 (1 school)	Small town approximately 40 miles from Pittsburgh	30%
F	1 (1 school)	Parochial school in Chicago	10%

Note: Low-income based on free and reduced lunch eligibility.

Tests and interviews were conducted by researchers and trained research assistants from Northwestern University. At first grade, whole-class tests were read aloud to the entire class, one question at a time, while students filled in their test booklets. Students were given sufficient time on each item before going on to the next, and the whole-class tests took approximately one hour.

Individual interviews took place outside of the classroom. Again, the researcher presented the questions one at a time and notes were taken regarding the child's work and response. During interviews, researchers probed students for their mathematical thinking and solution procedures. For additional information on sampling, procedures, and tests, see Drueck, Fuson, Carroll and Bell (1995) or Drueck, Fuson, and Carroll (1999).

Tests and Interviews

To provide for comparisons, 44 items on tests and interviews were selected from a recent international study of mathematical achievement of Japanese, Chinese, and U.S. students enrolled in a traditional mathematics program (Stigler, Lee, & Stevenson, 1990). The sample of U.S. schools from the Stigler et al. study (urban and suburban schools in the Chicago-area) was similar to that selected for the longitudinal study. Because the original study had tested the three international samples at first and fifth grade, it provided good comparisons for the first and last years of the longitudinal study. Furthermore, it allowed a comparison to Asian students in high-achieving nations, as well as to U.S. students in more traditional curricula.

In addition to these 44 questions, a one-minute addition fact test from the Stigler study was also used. An additional 27 questions investigating students' progress in other areas, including topics specific to EM, were included on the spring test. Note that for these 27 items there is no comparison group.

Results

International Comparison

On the one-minute addition test, EM first graders correctly answered 8.7 facts per minute compared to 8.2 for the Chinese, 8.9 for the U.S. comparison sample, and 11 facts per minute for the Japanese students. Thus, while the EM students performed substantially below the Japanese students, there was little difference from the other two samples.

On the 44 comparison items (11 on the fall test, the remaining 33 on the spring test and interview), EM first graders performed higher than both the Chinese and the U.S. comparison students, but below the Japanese students (Table 2.2). Compared to U.S. performance in other international studies, EM children performed very favorably relative to their Asian peers.

Table 2.2: Percent correct for each sample on 44 questions

Grade	EM	U.S. comparison	Japanese	Chinese
First grade	58%	43%	64%	52%

Notes: The sample sizes were Japanese, 750; Chinese, 1037; U.S. comparison, 976; EM, 503 on tests, 240 on interviews. See Table 2.3 for fall-to-spring comparisons.

Scores on each of the 44 comparison questions were tested for significant differences between the EM students and the other three samples. Chi-square tests at the .01 level ($\chi^2 \geq 6.64$) were used to test for statistically significant differences[3]. Table 2.3 shows the number of these items on which EM students scored below (>), equivalent to (=), or above (<) the Japanese, Chinese, or U.S. traditional students on each items. For individual items and scores, see Appendix A.1.

[3] Because of the number of items being compared, a more conservative .01 level of significance is used for most comparisons in the longitudinal study.

Looking at subtests of the Stigler items (Table 2.3), the EM students performed strongest relative to their peers on Number concepts and Word problems. For example, 23% of the EM students correctly identified the fraction of a circle shaded (1/4) compared to only 2% of the Japanese, 11% of the Chinese, and 6% of the U.S. comparison students. EM students also performed higher than all of the groups on the questions, “Write 10 more than 57” and “How can you arrange 3, 6, and 1 to form the biggest number?” Thirty-seven percent of the EM students correctly answered “Give the number that is the same as ten tens (100)” compared to 17% of the Japanese students. The EM students did significantly better than Chinese students on three of the five questions involving equations (e.g., 4 + 6 = ___ + 4). Across the board, the EM students outperformed the U.S. students in a traditional curriculum, and most of these differences were significant.

Table 2.3: Number of items on which the Japanese, Chinese, and traditional U.S. students did better than, worse than, or the same as the EM students

	Japanese			Chinese			Traditional U.S.		
Items	>	=	<	>	=	<	>	=	<
Fall whole-class test	8	2	1	4	5	2	1	2	8
Spring test total	10	20	3	1	21	11	0	9	24
Word problems	3	8	0	0	10	1	0	2	9
Number concepts	1	4	3	0	3	5	0	0	8
Equations	2	3	0	0	2	3	0	2	3
Estimation	1	3	0	0	3	1	0	2	2
Graphing	3	2	0	1	3	1	0	3	2

Note: All longitudinal fall test items were whole-class. Spring items were from both whole-class tests and individual interviews. The >, =, < compare performance of the named sample to the EM sample. For example, in the fall whole-class test (top left), the Japanese did significantly better than the EM sample on 8 questions, equal to the EM sample on 2 questions, and significantly worse than the EM sample on 1 question. See Appendix A.1 for individual items.

Between fall and spring, EM students showed a gain relative to each of the groups. For example, in the fall they performed lower than the

Chinese students on 36% of the questions and better on only 18%. In the spring, they performed below the Chinese students on only 3% and better on 33% of the questions. This fall-to-spring differential pattern, favoring the EM students, is similar relative to each of the groups, indicating a positive effect of the curriculum.

Many recent international studies of mathematical achievement, including Stigler's, show that U.S. students lag behind their peers in other nations. However, the first year of the longitudinal study provides evidence that this "learning gap" (Stevenson & Stigler, 1992) can be closed when children are enrolled in a more ambitious and challenging mathematics curriculum, such as EM. Children in EM are asked to investigate number concepts, to develop their own methods for solving problems, and to discuss these methods. They are asked questions that facilitate development of place value and number sense, such as "What is 10 more than 57?" and story problems are a regular part of the daily mathematics lesson. Topics that have been delayed or ignored in elementary school mathematics, such as geometry, fractions, and non-standard equations (e.g., 3 + 4 = 4 + ___) are a regular part of the EM curriculum. Thus, it is not unexpected that EM students performed well relative to the Asian students and much higher than comparison U.S. students on these types of questions.

Additional Test Questions

Along with the international comparison items, a number of other questions were included on the spring tests and interviews (see Table 2.4). These are divided into Number and Patterns, and Other. Results indicate that EM first graders have a good grasp of a wide range of mathematical topics.

Table 2.4 Results on spring interviews and tests: Percent answering each item correctly

Question		Percent correct
A.	NUMBER CONCEPTS	
1.	Place-value task	
	a. Reads the number "16."	97
	b. Counts out 16 chips.	95
	c. Tells what the "6 part" is. (6 ones)	93
	d. Tells what the "1 part" is. (1 ten)	23
2.	How many teams of 10 children could you make with these 53 children?	35
3.	(4 dimes, 5 nickels, 2 pennies in random order)	
	a. How many cents is this?	52
	b. How much more money do you need to make $1.00?	29
4.	(Picture of 5 bunches of grapes and 4 single grapes) Each bunch has 10 grapes. Then there are some extra grapes. Write the number that tells how many grapes there are altogether.	59
5.	Jan has 63 crayons. He put 10 crayons in each box.	
	a. How many boxes did he fill?	32
	b. How many crayons were left over?	46
6.	Completes partially filled in hundreds grid.	58
B.	PATTERNS AND NUMERATION	
7.	Fill in the missing number pattern: 6, 16, 26,__, ___, ___, ___	69
8.	Fill in the missing shapes to keep the pattern going: (t)riangle, (c)ircle, c, (s)quare, s, t, c, c, s, s, ___, ___, ___, ___, ___	91
9.	Write the numbers that keep the patterns going. 100, 96, 92, ___, ___, ___, ___	9
10.	Count the tallies and write the number (47 grouped by 5's)	64
11.	Single rule Frame and Arrow puzzle	
	a. Identifies correct rule (+2)	44
	b. Finds first missing term (9)	59
12.	Double rule Frame and Arrow puzzle	
	a. Identifies first rule (+3)	62
	b. Identifies second rule (– 6)	7
	c. Finds first missing term (12)	67
C.	OTHER PROBLEMS	
13.	Is the number even or odd?	
	a. 43	69
	b. 78	65
14.	a. Divide the square into thirds.	58
	b. Label the parts (1/3 or one-third)	59
15.	How far is it from 36 to 52?	34
16.	Joe was born on November 1, 1986. Sue was born on March 1, 1986.	
	a. Who is older, Joe or Sue?	56
	b. How much older?	15

Note: Items 1, 2, and 3 were given in individual interviews (n= 230). Other problems were from whole-class test, n = 503. On Problems 7–9, repeating patterns as well as continued patterns were accepted. For example, either 6, 16, 26, 36, 46, 56, or 6, 16, 26, 6, 16, 26 were accepted.

Chapter 3
Second Grade Study (1994–1995)

Participants and Procedure

At second grade, whole-class tests were administered to all students in 22 classes, 343 of whom were in the longitudinal sample. For the analysis, only tests from students in the first grade sample were used; the others were discarded. During May, individual interviews were also administered to 131 students who had been interviewed at the end of first grade.

Whole-class tests were developed and administered by researchers from Northwestern University. Each question was read orally as students followed along in their test booklets. Sufficient time was allowed for students to complete each question, with test administration taking approximately 60 minutes.

For comparative purposes, 24 questions from a recent study of U.S. and Japanese second graders (Okamoto, Miura, & Tajika, 1995; Okamoto, Miura, Tajika, & Takeuchi, 1995) were selected for the whole-class longitudinal test. Ten of these items were taken from Okamoto et al.'s number-sense test and 14 from their achievement test. Okamoto et al.'s study included U.S. second graders from a middle to upper-middle-class school in California using a more traditional mathematics program and Japanese second graders attending a middle-class school in Tokyo. Although the EM longitudinal sample was more heterogeneous (and hence might be expected to perform lower than Okamoto's samples of higher socioeconomic status), these test items provided an international comparison that built upon the first grade comparison. Chi-square tests at a .01 level of significance ($\chi^2(1) \geq 6.64$) were used to test for differences between EM students and the other samples.

Along with these 24 comparison items, additional longitudinal test questions investigated a variety of mathematical areas, including questions repeated from the first-grade test. For additional information on sample and procedures, see Fuson, Carroll, and Drueck, 2000.

Interview questions allowed researchers to investigate the problem-solving strategies employed by the students. The one-minute addition fact test from first grade was repeated.

Results

Whole-Class Tests

EM students had a higher mean score than either of the other two groups on the number sense test and a higher score than the U.S. sample on the achievement test (see Table 3.1). However, there were few significant differences on individual items on the number sense test. On the mathematics achievement subtest, EM students scored higher than the U.S. sample on four items (completing a pattern and symbolic addition and subtraction), but significantly lower than the Japanese students on half of the mathematics achievement items.

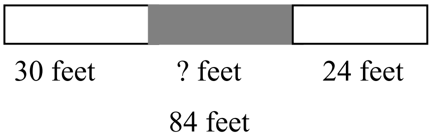
Results provide further evidence of the development of number sense in EM students. For example, EM students performed higher relative to the other groups on questions like “What number comes 9 after 999?” or “What is the smallest 5-digit number?” EM students were also more likely to do better than the other groups on computation problems that could be solved by knowledge of the number system, counting, or other alternative methods (e.g., 80 – 7 or 600 + 100). However, they did not perform as well on problems that might best be solved by algorithmic knowledge (536 – 127). Although the Japanese students scored lowest on the number sense test, they scored much higher than both the EM and U.S. comparison samples on the mathematics achievement test. These results on the achievement test mirror those on the first grade international comparison, with EM students scoring between the Japanese students and U.S. comparison students. As in first grade, EM students showed a strong number sense relative to the other groups.

Table 3.1: Percent correct on comparison questions

Item	EM n = 343	U.S. sample n = 29	Japanese sample n = 30
NUMBER SENSE TEST			
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
NUMBER SENSE TEST MEAN CORRECT	54	52	50
MATHEMATICS ACHIEVEMENT TEST			
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 *

(Table continues)

Table 3.1 continued

12. How long is the shaded area? 	24	10	56
MATHEMATICS ACHIEVEMENT TEST MEAN CORRECT	70	55	92

Notes: U.S. and Japanese samples are from Okamoto, Miura, Tajika, 1995.
The three parts of Question 1 on the mathematics Achievement Test were scored separately, so there were 14 questions in all.
* Difference from EM score significant at .01 or less on Chi-square test.

Although EM students scored lower than both of the comparison groups on the standard 3-digit computation from Okamoto et al. (see Table 3.1), this may have been because Okamoto's samples were from higher social-economic backgrounds than the more heterogeneous EM longitudinal sample. To investigate this issue further, EM scores on 2-digit addition and subtraction from the second-grade test (see Table 3.2) were compared to similar problems from a national standardized test (The Stanford Test, The Psychological Corporation, 1992). In this comparison, the EM students were above the spring second-grade national norms for 2-digit addition (65% for EM and 50% for U.S. second graders) and at the norm for 2-digit subtraction (38%).

Additional test items, including four follow-up questions from the first grade study, were also included to assess the strengths and weakness of EM students (see Table 3.2). Results provide further evidence that EM students have a very strong understanding and use of place value (Table 3.2). For example, 76% correctly wrote the number "five thousand four," 71% correctly wrote "100 less than 465," and 66% correctly wrote "6 tens, 3 ones, and 5 hundreds." EM students also scored high on fraction questions.

Table 3.2: Percent correct on additional items from second-grade test

Question	EM n = 343
Place value	
1. Write the number "five thousand four." (5004)	76
2. Write the number "three hundred twenty-six." (326)	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 (563)	66
6. Write the number that has 7 thousands, 8 tens, 5 ten thousands, 1 one, 0 hundreds. (57,081)	38
7. What is the number that is the same as ten tens? (100)	62
8. Complete the number grid (number grid with spaces not filled in).	84
Computation	
9. 36 + 47 (vertical format)	65
10. 72 – 26 (vertical format)	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. The 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

Four items from the first grade longitudinal test were repeated on the second-grade test. EM students made strong gains on all of these, as shown in Table 3.3.

Table 3.3: Percent correct on items from first and second-grade whole-class tests

Question	First grade n = 503	Second grade n = 343
Write the number that is 10 more than 57	68	85
Write the number that is the same as ten tens.	37	65
Circle half of the stars	32	65
Jim had 63 crayons. He put 10 in each box.		
a. How many boxes did he fill?	32	56
b. How many crayons were left over?	46	65

Individual Interviews

On the one-minute addition fact test, EM students had a mean correct score of 16.2 facts in one minute compared to 8.7 at first grade. Students were also asked to solve two addition and two subtraction fact problems, and the solution methods were categorized by the observer. When necessary, the student was asked to clarify his or her solution method. Table 3.4 shows the mean correct on each of the fact problems, and the main strategies used. While the EM curriculum suggests that students should achieve automaticity with the addition and subtraction facts by second grade, an analysis of strategies indicates that few students used direct fact knowledge (fewer than 25% on any of the problems). Instead, most used some type of counting (counting on from the most common) or a fact strategy (e.g., a derived fact). The mean correct rate was high (greater than 90%) on all but Item 4.

Table 3.4: Percent correct on facts and percent using each method

Fact	Mean correct n = 131	Percent using strategy (and % correct using this strategy)		
1. 7 + 9	.92	Direct fact	.10	(.92)
		Counting	.52	(.88)
		Fact strategy	.37	(1.00)
2. 6 + 7	.93	Direct fact	.13	(.85)
		Counting	.50	(.80)
		Fact strategy	.41	(.98)
3. 9 – 4	.93	Direct fact	.25	(.94)
		Counting	.53	(.93)
4. 14 – 9	.74	Direct fact	.16	(.81)
		Counting	.56	(.77)

Note: Percents using strategies in column 3 do not add up to 100% because some were classified as "other."

All students were also asked to solve 2-digit addition and subtraction word problems; students who were successful on these were then given a 3-digit addition or subtraction problem. Mean correct, patterns of errors, and strategies (e.g., mental computation, written algorithm) were recorded (Table 3.5). Only the most common strategies are included in the table.

Some other strategies included use of pictures, a number grid, or base-10 blocks, all of which were available for the student.

Some form of mental computation was most commonly used on 2-digit addition and subtraction. However, on 3-digit computation, more than half of the students used the standard written algorithms, most likely because of the load on memory imposed by the larger numbers. Results indicate that students were generally successful at solving addition problems, but less successful with subtraction, and students' level of performance on these word problems was nearly identical to that on similar computation problems on the whole-class test.

Table 3.5: Percent correct on word problems and percent using each method

Problem	Percent correct	Percent using strategy (% correct)		
1. Jack has 39 baseball cards. Mary has 26 cards. How many cards to they have altogether? (39 + 26)	69	Written algorithm Mental computation Fingers or tallies	33 56 06	(84) (64) (63)
2. There were 43 stickers in Ann's sticker book. She took out 26 stickers. How many stickers were left in the book? (43 – 26)	35	Written algorithm Mental computation Fingers or tallies	41 47 8	(39) (36) (20)
3. 369 cars were in the parking lot this morning. 174 more came this afternoon. How many cars were there altogether? (369 + 174)	56	Written algorithm Mental computation	56 33	(73) (43)
4. Sue had 512 M&M's. She ate 246 of them after school. How many M&M's did she have left? (512 – 246)	28	Written algorithm Mental computation	67 24	(35) (18)
5. One day there were some apples under a tree. The next day 6 more apples fell to the ground. Now there are 14 apples on the ground. How many apples were on the ground the first day. (___ + 6 = 14)	47	Written algorithm Mental math Fingers or tallies	2 83 8	(67) (47) (82)

Note: All 131 students were asked Questions 1, 2, and 5. Ninety-one students were asked Question 3 and 46 were asked Question 4.

Chapter 4

Third Grade Study (1995–1996)

Participants and Procedure

By third grade, the longitudinal sample had dispersed to 29 classes. Whole-class tests were administered to 620 students, 236 of whom had been in the original sample. Only the results of students from the original longitudinal sample were analyzed. Additionally, 138 of the students in the interview sample were individually interviewed. Tests and interviews were administered in May and June by researchers and trained research assistants from Northwestern University.

For whole-class tests, comparison items were taken from two previous studies. Twenty-two items from the Fourth National Assessment of Educational Progress (Lindquist, 1989) administered to a random sample of U.S. third graders were selected. Nine items were also taken from a third-grade cognitively-based test (CBT). Items from the CBT had been administered to third-graders in traditional instruction as well as those in a constructivist reform-based program, the Problem-Centered Mathematics Program (Wood and Cobb, 1989).

In order to include more items, four different forms of the whole-class test were constructed, each consisting of 33 questions. In administering the test, the first items were taken from the CBT and read orally (as in the original Wood and Cobb study). The rest of the test, including the NAEP items, was administered in written form. The last problems on each test were performance-based items requiring a ruler or other tool. When students reached these items, they were given the necessary tools. Test administration took approximately 50 minutes.

Chi-square tests were used to compare EM students to the comparison samples on both the NAEP and CBT. For these comparisons, a .01 level of significance was used, $\chi^2(1) \geq 6.64$. For additional information on sample and procedures, see Fuson, Carroll, and Drueck, 2000.

Results

Whole-class Test

NAEP comparison items On the 22 NAEP questions, the EM students had a mean correct score of 61% compared to 44% for the NAEP sample (see Table 4.1). For further analysis, the questions were classified into two subtests, number concepts and computation and geometry, data, and reasoning, each containing 11 questions. On the number concepts and computation subtest, EM students had a mean percent correct of 65% versus 52% for the NAEP sample. On the geometry, data, and reasoning subtest, EM students had a mean percent correct of 56%, compared to 35% for the NAEP sample. The EM sample scored significantly better than the NAEP sample on six of the number items and eight of the geometry, data, and reasoning items. Groups were equivalent on the remaining eight items.

Because there is some concern about the computational skills of EM students (and more generally, students in *Standards*-based curricula, where rote computation is given less attention), the results in this area are of special interest. That is, because less time is spent on drill and massed practice in a curriculum like EM, are the students weaker at computation? On five of the symbolic computation items, there were no significant differences between the groups, while the EM students scored higher on Item 8, which assessed the relationship between addition and subtraction. On all story and place-value questions, EM students significantly outperformed the NAEP sample.

Table 4.1: Percent correct on third-grade NAEP comparison items

Question	Everyday Math n = 107 to 236	4th NAEP n = 18,033
NUMBER CONCEPT AND COMPUTATION		
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 *
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)	56	29 *
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 cost 30 cents, a roll of tape costs 35 cents, and box of nails costs 20 cents. What is the cost of a roll of tape and a package of screws?	77	58 *
NUMBER SUBTEST MEAN	65	52
GEOMETRY, DATA, AND REASONING		
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 4-by-7 rectangle?	23	15 *
b. With length and width shown, 4 by 7	67	17 *
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
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 *
GEOMETRY, DATA, AND REASONING MEAN	56	35
OVERALL MEAN	61	44

Notes: For details on sampling, see Fuson, Carroll, and Drueck. * p = .01 or less on Chi-square test.

Results on the nine items from the CBT provide further evidence that EM students perform better than their peers when computation is presented in a problem-solving context. EM students scored significantly higher on all story problems (Table 4.2). EM students also scored higher on one of the place-value questions as well as on the multiplication-division fact item.

Table 4.2: Percent correct from cognitively-based test (Wood and Cobb)

Question	Everyday Math n = 236	Comparison sample n = 191
A. 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 arrive. 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 *
B. 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 and 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-logs and 2 unit-cubes (base10 blocks), and a box.)	73	50 *
C. Multiplication and division		
3. 3 * ___ = 27	80	59 *

Notes: Because different forms of the test were given, the number of EM students varied from a total sample of 236 across sub-samples of 107, 117, and 119.
* p = .01 or less on Chi-square test.
Cobb's sample included students in traditional instruction and the Problem-centered Mathematics Program.

Comparison of results between the longitudinal second and third grade tests indicate that EM students made good progress in multi-digit computation (Table 4.3).

Table 4.3 Second and third-grade comparisons

Question	EM End of second grade	EM End of third grade
1. 80 – 7	67	82
2. 110 – 40	50	80
3. 296 + 604	54	78
4. 72 – 26	38	
54 – 37		72

Performance-based Items Various studies (Kenney and Kouba, 1997; Lindquist and Kouba,1989; Peak, 1996) have shown that U.S. students are weak in their knowledge of geometry and measurement, especially when questions require more than simple identification. In contrast to traditional programs, students in EM typically take part in hands-on investigations, including projects that involve measurement, construction, and investigations of more complex geometric figures. To assess their knowledge in this area, 11 performance-based questions were included on the third-grade longitudinal test. These questions required students to use a ruler or the geometric template to draw and measure figures.

As results in Table 4.4 show, EM students are skillful at measurement involving both inches and centimeters, even when fractions of the units are involved. For example, between half and two-thirds of the students could draw or measure line segments, and about 40% could construct figures with given measurements. These strong results show the progress that elementary students can make when geometry and measurement are investigated in greater depth.

Table 4.4 Percent correct on performance-based items

Question	Percent correct
1. Drawing figures	
Draw a quadrangle.	64
Use your template to draw a parallelogram.	41
Use your template to draw a hexagon.	80
Divide the hexagon into 3 equal parts.	26

2. Measurement	
Use your ruler to measure the line segment:	
In inches (4 ½ in.)	49
In centimeters (12 cm)	58
Use your ruler to draw a line segment that is:	
2 ½ inches long.	62
10 cm long.	74
Draw a square that is 6 cm × 6 cm.	37
Draw a rectangle that is 4 cm × 8 cm.	43

Notes: Each question answered by about half of the students. Measurements were correct if within ¼ inch or ½ cm of the correct measure; squares and rectangles required both correct measurement and right angles.

Interviews As in second grade, children were asked to solve story problems involving addition and subtraction. At third grade, 76% of the students interviewed correctly solved the 3-digit addition problem and 52% correctly solved the 3-digit subtraction[4]. These results show strong gains over performance at third grade where only 35% and 28% correctly answered 3-digit addition and subtraction. On both of these problems, most third graders used the standard written algorithms (88% and 95% respectively), a higher percent than at second grade. Given the level of difficulty of the problems (3-digits requiring regrouping), the use of the standard algorithms might indicate a good mental flexibility, i.e., recognizing which procedure might be less error prone for the problem[5]. It also may reflect the presentation of the question since, during interviews, students were provided with paper and pencil, perhaps cueing them to use a written procedure.

Despite the increased use of the standard algorithms on addition and subtraction, the students interviewed showed a continued familiarity with

[4] The addition problem was "On Saturday, 264 fans came to the first baseball game. 379 fans came to the second game. How many fans came to the two baseball games?" The subtraction problem asked: John collected 512 ants for his ant farm. Bill dropped the ant farm and 346 ants escaped. How many ants does John have left?"

[5] Fourth grade interviews included addition and subtraction problems that were better fitted to mental strategies, and students were very successful at these.

alternate solution algorithms. For example, students were shown the problem 63 – 19 and told that a student solved it by first subtracting 20 and then adding 1 more to the answer (this solution procedure was illustrated for the student on a card). When asked if this method was suitable, 78% said that it was. When then asked to solve 53 – 28 by this method, 73% did so correctly.

Children were also asked to use the calculator to solve a subtraction and a division problem. Seventy percent of the children correctly solved the problem 20,748 – 8,967, and 93% correctly solved 1978/23.

Summary of Third-Grade Results

EM students made strong gains in multi-digit addition and subtraction between second and third grade. Items from the NAEP as well as from a cognitively-based mathematics test showed that EM students scored the same as comparison students on symbolic computation and much higher in all other areas, including computation in story problems. Scores on performance-based items were also very high compared to students in traditional curricula as evidenced by results on the National Assessment of Educational Progress (Kenney and Kouba, 1997; Lindquist and Kouba, 1989).

Overall, results indicate continued progress in all areas of mathematics, familiarity with mathematical tools, knowledge of alternate solution methods, and a good ability to apply mathematics in problem-based situations.

Chapter 5
Fourth Grade Study (1996–1997)

Participants and Procedures

Students from the longitudinal study were tested and interviewed at the end of fourth grade. Of those who were administered the whole-class test, 170 were from the original longitudinal sample[6]. One hundred four students were also individually interviewed.

Of the schools in the study, seven of the schools were in the Chicago area and the remaining one was a small city school in Pennsylvania. Districts were urban, suburban, and small city/rural and included students from a wide range of racial and economic backgrounds, including one Spanish bilingual class in Chicago.

Two forms of the whole-class test were constructed, each consisting of 37 questions. Because there was a wide range of social-economic status in the schools, a block design was used to assign classes to the test forms, ensuring that students from districts with both higher and lower SES took each form of the test. To allow comparisons, 30 fourth-grade problems were selected from the 1990 and 1994 NAEP (Educational Testing Services, 1990, 1992). These NAEP items assessed knowledge in five areas: numbers and operation, algebra, geometry, measurement, and analysis and graphing. On each of the 30 NAEP items, differences between the EM and NAEP samples were tested with a Chi-square test, $\chi^2(1) \geq 6.64$, $p < .01$.

Additional test questions included 11 performance-based items that required use of a measurement or geometry tool. Two forms of the individual interview were also constructed assessing student's knowledge and strategies in mental computation, estimation, place-value knowledge, and geometry. Interviews began with a one-minute multiplication facts test.

Tests and interviews were administered to whole classes during April and May by researchers from Northwestern University. Directions were read aloud, and students worked independently without calculators or other tools. However, because the final items were performance-based geometry and measurement items, a ruler or geometric template was made available when students reached this point on the test. Tests and interviews each took approximately 50 minutes.

Results

Whole-class Tests

On the 30 NAEP items, EM students had a mean correct score of 71% versus 44% correct for the NAEP sample. The differences between the groups were largest on the algebra, geometry, and measurement subtests, and smallest on the number and operations and the data analysis subtests (see Table 5.1). Because computation has been a concern of EM teachers, the computation, story problem, and place-value scores are examined separately. Although χ^2-square tests showed no significant differences on the two symbolic subtraction problems, the EM students scored significantly higher than the NAEP sample on all of the number stories and on 5/6 of the place-value items. On individual items, EM scored significantly higher than the NAEP sample on 23 of the items and lower on none. See Appendix A.2 for individual items and results.

Differences were especially strong on the three performance-based items: drawing a square with two vertices given (EM 91%, NAEP 40%);

---

[6] Two schools, both public schools in Chicago, used the EM curriculum only through third grade, and these students were no longer included in the longitudinal sample. Additional students were lost through attrition

drawing a rectangle with dimensions 2 inches and 3 ½ inches (EM 67%, NAEP 18%); drawing a rectangle of a given area (EM 80%, NAEP 43%).

Table 5.1: NAEP comparison—percent correct on subtests

Subtest (number of items)	4th grade EM	4th grade NAEP sample
MEAN ON NUMBER AND OPERATIONS (15)	72	49
Computation	62	58
Story problems	65	36
Place value	84	61
MEAN ON ALGEBRA (3)	62	35
MEAN ON GEOMETRY (3)	75	34
MEAN ON MEASUREMENT (5)	71	39
MEAN ON DATA ANALYSIS (3)	67	53
OVERALL MEAN ON ALL 29 NEAP ITEMS	71	44

Individual items and the results of the EM students and the NAEP comparison sample can be found in Appendix A.2.

Table 5.2 shows percent correct on all of the performance-based questions on the test. Items 1, 2, 3, and 4 were repeated from third grade, and students showed gains of 22%, 10%, 24%, and 38% respectively. Most of these questions were correctly answered by two-thirds or more of the students—very high rates for performance-based questions.

Table 5.2: Percent correct on performance-based questions

Question	Percent correct
1. Measure line segment AB to the nearest centimeter. (Line segment 8 cm)	80
2. Use the ruler on your template. Draw a line segment that is 2 1/2 inches long.	72
3. In the space below, draw a rectangle 2 inches wide and 3 ½ inches long.	67
4. Draw a square that is 6 cm by 6 cm.	75
5. On the grid below, draw a rectangle with an area of 12 square units.	80
6. On the grid below, draw a rectangle with a perimeter of 14 units.	21
7. Use your template to draw a square with two of its corners at the points below (2 vertices given).	91

Table continues

or absence on the days of testing.

8.	At 2:20, the hands of the clock form an acute angle. Draw a time when the clock's hands form an obtuse angle. Write this below the clock.	
	Correct angle drawn	64
	Time matches time shown	82
9.	Copy a figure from your template that matches the clues to the mystery figure: "*I am a quadrangle. Both pairs of opposite sides are parallel. I have no right angles*."	56
10.	If the large hexagon on your template is 1 whole, then what shape on your template would represent 1/3? Copy this shape from your template.	65
11.	If the large hexagon on your template is 1 whole, then what shape on your template would represent 1/6? Copy this shape from your template.	61

Comparisons on various symbolic computation problems included on both the longitudinal third and fourth tests also showed good gains. The EM students are now near ceiling (94% correct) on 3-digit addition requiring regrouping, and about two-thirds correctly solved 3-digit subtraction requiring regrouping. Half or more of the students correctly solved multi-digit multiplication and division problems. Comparisons to U.S. students show that EM students were performing equivalent to 5th graders in traditional curricula on these problems (Stigler, Lee, and Stevenson, 1990).

Table 5.3: Percent correct on multi-digit computation

Problem	EM 3rd grade	EM 4th grade
49 + 56 + 62 + 88	60	76
296 + 604	78	
296 + 304		94
504 – 306	38	
604 – 207		56
504 – 306 (horizontal)		65
242 – 178		69
242 – 178 (horizontal)		68
198 × 4		77
45 × 56		52
$6\overline{)432}$		49

Note: Problems were in vertical format unless otherwise noted.

Individual Interviews

On the one-minute multiplication fact test, students attempted a mean of 23.6 questions and had a mean correct score of 22.6. This translates to 2.6 seconds per fact, which is within the 3-second rule sometimes used to assess automaticity of facts.

One of the goals of the EM curriculum is that students will develop a flexible number sense, including facility with a variety of computational strategies such as counting up by tens and ones or changing numbers to "friendlier" numbers, e.g., 147 – 99 = 147 – 100 + 1. While many third graders used the standard written algorithms on the interviews, the third-grade interview problems were not well suited for mental computation. On the fourth-grade interviews, students were given two problems and asked to solve them mentally and then explain their method.

As shown in Table 5.4, only about one-third of the students used the standard written algorithm mentally. More typically, students changed the problems to friendlier numbers (e.g., 399 + 26 + 30 = 400 + 25 + 30) or used another mental methods such as adding hundreds, then tens, and then ones. In short, more than one-third recognized that the calculations could be simplified by changing the numbers, and this group had the highest correct rate. Less than one-third used the standard algorithms mentally, i.e., "borrowing" or "carrying" in a right-to-left direction.

Table 5.4: Percent using each mental strategy (percent correct using that strategy)

Problem	Made into "friendlier" numbers for mental computation	Borrowed or carried mentally	Other mental
399 + 26 + 30	35 (95)	31 (53)	30 (63)
425 + 50 + 25	40 (95)	34 (53)	26 (53)

Students were also asked to estimate the solution to the problems 57 * 11, 4023 – 37, and 512 – 189 and to explain their methods. Most children gave estimates within an acceptable range, with few students calculating the exact answer and then rounding. Instead, students used a variety of reasonable estimation techniques.

Students were also asked to identify the value of the underlined digit "2" in four different numbers. The numbers and the percent answering correctly for each are 210 (94%), 12.5 (46%), 5,200,610 (74%), and 409.25 (63%).

Summary of Fourth-grade Study

The EM fourth-grade sample scored much better than comparison U.S. students on the NAEP items, with especially large differences on algebra, geometry, and measurement. These results are not unexpected, given the increased time spent in EM on these topics from kindergarten onward. A number of educators and researchers have explained the low scores of U.S. students on international comparisons as a result of *decreased opportunity to learn* advanced topics relative to their counterparts in Japan, Germany, and other industrialized nations (Peak, 1996). The EM curriculum has increased the opportunity to learn algebra, geometry, and other topics often ignored in elementary school. Results on performance-based items shows that EM students continue to make good progress on these more complex tasks.

Test results also show continued progress in solving symbolic addition and subtraction problems, while interviews indicate that EM students have a good number sense, as indicated by their estimation and mental computation as well as an awareness of multiple solution methods.

Many students used sophisticated mental strategies to solve multi-digit computation.

Chapter 6
Fifth Grade Study (1997–1998)

Participants and Procedures

One hundred seventy-one fifth graders from the longitudinal sample were tested and 89 were interviewed at the end of the school year (late April through May). Whole-class tests and individual interviews were administered by researchers from Northwestern University and took approximately 50 minutes.

To increase the number of items while holding down the testing time, two forms were constructed for both whole-class tests and interviews. Thirty-three end-of-year questions were selected from the Stigler, Lee, and Stevenson (1990) international study of mathematical achievement (the same study that provided the comparison sample used in the first year of the longitudinal study). This provided both an international comparison (Japanese, Chinese, and U.S. students in a traditional program) as well as a longitudinal comparison (the progress of EM students from first to fifth grade relative to these groups).

Twenty of these Stigler comparison items were on whole-class tests and 13 on individual interviews. Questions assessed a wide range of topics, from mental computation to geometry and algebraic thinking. Differences between the EM and each of the samples were tested with a Chi-square test, $\chi^2(1) \geq 6.64$, $p < .01$. (Six additional computation items were selected from the Stigler study, but these had been given to the comparison samples in the fall and so are discussed separately.)

For additional comparison, seven items were selected from a recent study of mental computation (Reys, Reys, & Hope, 1993). The sample from the Reys study included fifth-graders from U.S. suburban and Canadian urban districts. One division word problem was also chosen from a study by Silver, Shapiro, and Deutsch (1993) to assess understanding of long division and the meaning of the remainder. Selected items from previous years of the longitudinal study were also included on the longitudinal tests and interviews.

Whole-class tests began with a one-minute timed multiplication facts test. Following this, students worked independently without calculators or other tools until they reached the performance-based items at the end of the test. When students reached this point, they were provided with geometric templates or protractors.

Results

Comparison Items

International comparison On the 33 Stigler items, EM fifth graders had a mean score correct of 75% compared to 80% for the Japanese, 76% for the Chinese, and 44% for the U.S. fifth graders in traditional instruction. Thus, from first grade to fifth grade, the EM students have kept pace with the Japanese and Chinese students, while the U.S. comparison sample has fallen further behind (from 21% behind the Japanese students in first grade to 36% behind in fifth grade).

In a comparison of individual items, EM students showed improvement relative to the Japanese students and a decrease in performance relative to the Chinese students (Table 6.1). For example, in first grade, the EM students performed as well as the Japanese students on only 50% of the items and below them on 41%. By fifth grade, the EM students performed as well as the Japanese students on 82% of the questions and below them on only 18%.

Table 6.1: Percentage of questions on which fifth-grade EM students performed better than, the same, or worse than international samples

	Japanese			Chinese			U.S. Traditional		
Overall Percentage	>	=	<	>	=	<	>	=	<
First Grade	41	50	9	11	59	30	2	25	73
Fifth Grade	18	82	0	24	69	6	0	24	76

Note: Comparison samples are from Stigler, Lee, and Stevenson (1990). First grade tests included 44 comparison items, fifth grade tests 33 items. The >, =, < compare performance of the named sample to the EM sample. For example, at fifth grade, the Japanese fifth graders did better than the EM fifth graders on 18% of the questions, the same on 82%, and worse on none. Chi-square tests at the .01 level of significance were used to test differences.

Results on the Stigler fifth-grade items are reported by subtest in Table 6.2. EM students had a higher mean correct score than the Chinese students on both the Word problems and Number concepts subtests. For example, 96% of the EM students correctly read 20,350,076 compared to 76% of Chinese fifth graders; 86% of the EM students correctly formed the largest 5-digit number using 2, 6, 3, 5, and 1 compared to 77% of the Chinese students.

Table 6.2: Percent correct on subtests from the Stigler study

Subtest (number of items on subtest)	EM	Japanese	Chinese	U.S. comparison
Word problems (8)	69	70	62	39
Number concepts and equations (10)	88	92	80	59
Geometry (7)	63	80	86	19
Mental computation (8)	73	78	77	54
Overall mean correct	75	80	76	44

Note: See Appendix A.3 for results on individual items.

EM fifth graders scored lowest relative to the Asian students on the Geometry subset, with both the Japanese and Chinese fifth graders correctly answering greater than 20% more of the geometry questions (Table 6.2). However, the gap between the EM students and U.S. comparison sample was also greatest on this subtest (63% versus 19%). When asked to find the area of a rectangle labeled 24 × 8 or to find the measure of the missing angle of a triangle (given two angles), these questions were answered correctly by 69% and 38% of the EM students respectively, but by only 8% and 0% of their peers in more traditional instruction. Again, the opportunity to learn more geometry has been of benefit, although the scores of the Asian students show that still more progress can be made. Individual results on the 33 comparison items are shown in Appendix A.3.

Other Comparison Items

On the one-minute multiplication test, students attempted 28.7 facts and had a mean correct score of 27.3 facts, an increase of 4.7 facts per minute over fourth grade. This translates to 2.2 seconds per fact.

EM students significantly outperformed the comparison fifth graders (Reys, Reys, and Hope, 1993) on all mental computation items. On most of these items, EM students had a mean correct score more than double that of the comparison group (Table 6.3). Differences were especially strong on the multiplication and division items (Problems 3 and 5), as well as on a subtraction question and the story problem. Interviews showed that EM students were more likely to use mental strategies rather than applying the standard algorithms mentally. For example, on 325 + 25 + 75, seventy-five percent of the EM students used a strategy such as "make friendlier numbers" (325 + 100). On the story problem, students were more likely to use a strategy such as add up (e.g., 65 + __ = 100) than to "borrow" and subtract.

Interestingly, EM students were much more likely to use the strategies on these problems than they were on the Stigler mental computation items discussed above. For example, 69% of the EM students used the standard algorithm mentally to solve 418 + 376 from the Stigler mental computation test, while only 18% did so when solving 325 + 25 + 75. Results from third-grade interviews also found that students used the standard algorithms during the interviews. These results suggest that the EM students are more flexible in their solution methods, recognizing problems that can be transformed easily and applying a more algorithmic approach to other problems. However, more research is needed to investigate this further.

Table 6.3: Percent correct on mental computation items from Reys et al. Test

Question	EM n = 40	Comparison Fifth graders
1. 47 plus 29	85 *	35
2. Double 84	80 *	50
3. What is 60 multiplied by 70?	70 *	33
4. 325 + 25 + 75	68 *	39
5. 3500 divided by 35	65 *	16
6. 7000 – 4000 – 300	73 *	18
7. Chuck's family lives 100 km from Chicago. They stop after driving 65 km. How much farther do they have to go?	88 *	32

Note: * indicates a difference of .01 or greater on Chi-square test.

Knowledge of alternative solution methods was investigated further in interviews. Because two multiplication algorithms are presented in the EM curriculum, the lattice and partial products methods, students were shown examples of each along with the standard written multiplication algorithm (multiply and carry) and asked which methods they know, and which they preferred to use. (The three multiplication algorithms were illustrated on cards, each showing a worked-out problem.) Students were then asked to solve a multiplication problem by any method. As Table 6.4 shows, most students were familiar with all three multiplication algorithms, including the standard method that is not formally taught in the EM curriculum, and students used all three methods to solve the problem. Although the standard algorithm was used by the greatest number of students, it also produced the most errors, while nearly all who used the lattice method were successful.

Table 6.4 Percent recognizing, preferring, or using a multiplication algorithm

Question	Lattice	Partial Products	Standard multiplication
Which do you know?	78	90	95
Which do you like best?	30	43	25
Method used to solve 35 * 27 (percent of these correct)	30 (92)	23 (78)	40 (38)

Note: The lattice method was taught in *Third Grade Everyday Mathematics*, but not included in the *Fourth Grade Everyday Mathematics*.

The division word problem taken from the Silver et al. study (1993)[7] was answered correctly by 43% of the EM students. This was the same percent as sixth, seventh, and eighth graders in the original study. Silver further found that 10% of his students gave an answer that included the remainder (counted as incorrect), e.g., 13.5 buses. Nearly the same proportion of EM students, 9%, made a similar error. Thus, EM fifth graders performed as well as the comparison junior high school students on this long division item.

Additional comparison items were selected from the Stigler et al. study (1990). However, because these were presented in November in the original study and in the spring to the EM students, direct comparisons are not possible. Still, results in Table 6.5 suggest that EM students are on track with symbolic computation relative to their U.S. peers in more traditional programs. However, performance is considerably below that of Asian

[7] The problem was "The Clearview Little League is going to a Pirates' game. There are 540 people, including players, coaches, and parents. They will travel by bus, and each bus holds 40 people. How many buses will they need to get to the game?" This problem was given to sixth, seventh, and eighth graders in the original study.

students on most problems, and considerable progress is possible. Additional addition and subtraction items show that most students are proficient at multi-digit addition and subtraction involving regrouping (or some similar strategy).

Table 6.5: Percent correct on symbolic computation items

Question	EM fifth graders	Stigler U.S. sample
1. 5.3 – 4.6	70	41
2. 38.15 – 9.43	69	54
3. 198 × 4	73	73
4. 45 × 26	78	54
5. 3281 divided by 6	77	53
6. 3572 divided by 46	36	15
7. 1457 + 698	81	—
8. 242 – 178	78	—
9. 1937 – 950	64	—
10. 2000 – 125	80	—

Note: U.S. comparison group was asked Items 1–6 in November and EM students were asked them in April/May.

Performance-based Items

Interviews and tests both contained performance-based items and open-ended questions, some of which had been asked in previous years. Results on these questions are shown in Table 6.6, with fourth-grade scores when appropriate. While there have been increases in most areas, the geometry questions show the biggest gains, perhaps reflecting the increased emphasis on formal geometry in fourth and fifth grade EM. These results are especially strong given the high level of geometric thinking required. For example, in Question 4, students must coordinate the geometric properties given and recognize that it identifies a non-rectangular parallelogram, a task that would be classified as Level 1 or 2 in the van Hiele model of geometry (Carroll, 1998). It also requires rejecting figures like regular hexagons that fulfill part of the description (opposite sides parallel and no right angles) but not other properties (it is a quadrangle). While this level of thinking is more generally explored at the junior high level in U.S. schools (if at all), 75% of the EM students selected a correct figure from the geometric template.

Teachers familiar with the knowledge of junior high school students will probably recognize the high performance of EM students on finding the diameter and circumference of a circle, the use of the circumference formula, and use of protractors.

Table 6.6: Percent correct

Question	Fourth grade	Fifth-grade
Measure and geometry		
1. Draw a segment 2 ½ inches long.	75	82
2. Draw a rectangle that is 2 inches wide and 3 ½ inches long.	67	78
3. On the grid, draw a rectangle with an area of 12 square units.	81	78
4. Copy a correct figure from your template: I am a quadrangle. Both pairs of opposite sides are parallel. I have no right angles.	56	75
5. Use properties to identify a geometric figure that is partly hidden.	48	64
6. Measure the diameter of the circle shown.	—-	70
What is its radius?		63
What is its circumference? (Formula $C = \pi D$ given)		37
7. Use your protractor to	—	
Measure acute angle.		82
Draw an obtuse triangle.		78
Measure your obtuse angle.		66
ESTIMATION		
8. What is a good estimate of 57 × 11?	62	57
9. What is a good estimate of 512 – 189?	26	53
STATISTICS		
10. Tawanda checked the prices of a can of soda at 5 different stores. (prices given)		
What is the median price?	61	69
The maximum price?	89	98
The range in prices?	44	54

Summary of Fifth-grade Results

Stigler, Lee, and Stevenson's international study provided bookends for the longitudinal study, with comparisons to Japanese, Chinese, and U.S. students at first and fifth grade. Longitudinal results showed that while U.S. students in a traditional curriculum fell further behind the Asian samples, the EM students kept pace with both the Japanese and Chinese students over these four years. At first grade, mean scores were similar to the Asian samples and remained so at fifth grade. Furthermore, some gains were made, with EM students performing better relative to the Japanese students at fifth grade than at first grade. Because previous studies have found U.S. students losing ground to their peers in other nations as they progress through school, these findings suggest the gains that can be made in a more ambitious curriculum.

Relative to the Asian students, EM students scored lowest on the geometry questions. However, other results show that EM students have a strong understanding of measurement and geometry, especially on performance-based items. Still, the international comparisons indicate that further progress could be made by EM students in geometry and other areas.

EM students are especially strong at mental computation. Interviews indicate they typically possess more than one solution method and have mental flexibility, e.g., they know alternative methods and often choose the method that best fits the problem. This is contrary to the usual situation in which mathematics is typified by rote use of algorithms (Stodolsky, 1988). Results on fact knowledge and symbolic computation indicate that EM students are continuing to make good progress.

Chapter 7
Teacher Feedback and Class Observations

Surveys and Interviews

Teacher feedback was obtained through interviews (over the phone, at meetings, and during class visits) and surveys. At the primary grades, only teachers whose classes were part of the longitudinal study were interviewed and surveyed, while a broader sample of teachers was surveyed at the fourth and fifth grades.

Many of the interview and survey questions focused on topics specific to the EM curriculum, e.g., what changes should be made in the revision, what lessons were difficult to teach. This lesson-specific feedback was analyzed at each grade level and presented to the authors at the University of Chicago School Mathematics Project and the publisher of *Everyday Mathematics*, Everyday Learning Corporation.

Overall, feedback showed that teachers were very positive about most aspects of the curriculum, and this was consistent across grades. Positive aspects commonly noted were:

- The problem-solving focus of the program
- The inclusion of higher-level thinking skills for all students, e.g., the emphasis on reasoning and mathematical communication
- The use of everyday situations and applications in the activities
- The emphasis on conceptual understanding of mathematics
- The shift from only arithmetic to a broader range of mathematical topics, e.g., algebra, geometry, and estimation
- The hands-on approach in which manipulatives, tools (rulers and geometric templates), and mathematical representations (e.g., fraction bars or hundreds grids) are incorporated in the lessons and activities

- The integration of mathematics into other areas of the curriculum, such as science and social studies
- The use of small groups and discussion to facilitate mathematical exploration and understanding

At all grades, teachers also noted that children were more positive about mathematics and showed an increased enjoyment or enthusiasm over previous years. The use of small groups and games as well as invented procedures were noted as increasing students' positive perceptions of mathematics.

Many of these positive aspects of EM overlap. For example, conceptual understanding of mathematical topics is enhanced by the use of representations that illustrate or support the principle being explored, and better understanding is often obtained when students apply and discuss mathematics in situations that they can understand. Teachers often noted that children surprised them with the sophistication of their solution procedures and their mathematical thinking. A number of teachers said that they themselves had developed a better understanding of mathematics through teaching EM. Not only did the students appear to be more motivated by mathematics, but the teachers also seemed to enjoy it more.

Teachers identified some weaknesses in the curriculum. These were commonly mentioned:

- There is insufficient practice of certain skills, especially computation and facts.
- Lessons sometimes jumped around too much, especially at the lower grades. Lessons were sometimes packed with too many unconnected activities.

- While some teachers welcomed the inclusion of advanced concepts for lower-achieving children, others reported that their lower students struggled with some of the activities.
- Some teachers felt that students were weaker at arithmetic facts, and the games and other routines did not provide sufficient facts practice.
- Parents did not understand EM, e.g., why the standard algorithms were not being taught.

The most common concern brought up by teachers was the role of computation in the EM curriculum. While teachers supported the use of invented algorithms and multiple solutions, they also felt that some students were not as strong at computation as students had been in the past. Many of the teachers felt that the standard school algorithms should be included in the program, especially since standardized tests still included a computational subtest.

These concerns reflect one of the difficulties in implementing a reform-based curriculum while "old" assessment tools are still used. While numbers and operations are a major strand in the EM curriculum, symbolic computation has been given a decreased role. At the same time, tests like the Stanford Achievement Test and the Iowa Test of Basic Skills continue to include computation as a major component. Furthermore, many parents believe that facts and computation should be the focus of the elementary school program.

As the EM curriculum undergoes a revision, many of these concerns are being addressed. Balancing the development of rapid paper-and-pencil computation skills and a flexible number sense is not easy, nor is increasing the number of topics taught without overburdening teachers and students.

The authors are currently working to address these concerns in a second edition of EM.

Classroom Observations

Classes participating in the longitudinal study were observed and videotaped throughout the year, and these videotapes and transcripts were analyzed. At all grades, observations showed mathematical classrooms that differed markedly from traditional mathematics classrooms (Stodolsky, 1988). EM students engaged in a much wider range of mathematical topics, learning environments and styles, and used a much wider range of materials. In contrast to traditional programs, where individuals typically practiced sets of problems, EM students also worked in groups, participated in whole-class discussions, played mathematical games, took part in projects, and used a variety of manipulatives and materials. The mathematics was generally more varied and challenging (Carroll, Fuson, and Diamond, 2000).

Teachers were successful in implementing many of the reforms suggested by the EM curriculum and the NCTM *Standards* (1989, 1991, 1995). Students felt secure in exploring mathematical topics and discussing their thinking and solution methods, and teachers supported students' thinking and explanations (Fraivillig, Murphy, and Fuson 1999). For example, in observations of first-grade lessons that focused on addition and subtraction number stories, all teachers elicited a variety of number stories from their students and encouraged students to explain how they solved the problems (Carroll, Fuson, and Diamond, 2000). Teachers often modeled the stories, and children used a variety of representations to illustrate and solve the stories.

In a replication of a previous study (Fuson, Ding, & Perry, 1997), EM second-grade teachers were more likely to ask problem-solving and conceptual questions than either Japanese teachers or U.S. teachers using traditional textbooks. Researchers found that the major reason for this was the EM curriculum; that is, the teachers asked higher-level questions that were explicitly suggested or stated in lessons. Teachers asked their own problem-solving or conceptual questions much less frequently. Still, the higher performance than even Japanese teachers is especially impressive and indicates improvements that can be made via the curriculum.

Classroom observations indicated areas in which EM teachers might benefit from additional support. Based on her observations of first-grade EM classrooms, Fraivillig developed a framework for reform-based teaching (Fraivillig, Murphy, and Fuson, 1999; Fraivillig, 2001). In this framework, exemplary reform-based teaching is characterized by teachers who elicit, support, and extend students' mathematical thinking. Fraivillig found that most EM teachers supported students' thinking (e.g., helped students give their explanations), but fewer elicited multiple solutions, and very few extended this thinking (e.g., help other students to understand the student's explanation). However in an analysis of first-grade transcripts, Carroll (Carroll, Fuson, and Diamond, 2000) found that most EM teachers in his sample elicited multiple word problems and solutions methods from their classes.

Researchers at Northwestern University reported that most of the EM teachers had made the first steps towards *Standards*-based teaching, and the classroom environments differed greatly from traditional mathematics classrooms. Class observations also indicated that teachers would benefit from more support in the EM curriculum in order to make progress in more difficult areas, e.g., orchestrating discussions of alternatives solution methods. For example, teachers might benefit from some suggestions on how to model students' thinking with overhead base-ten blocks or the hundreds grid.

An analysis of lessons and class observations indicated that teachers would also benefit from a more streamlined curriculum in which lesson goals were more coherent to the teacher and student. This concern was also voiced by teachers at the various grades who felt that lessons often contained too many activities and goals. As with concerns about computation, this is being addressed in a forthcoming second edition of EM.

Despite some difficulties in fully implementing *Standards*-based teaching, observations showed these EM classrooms to be exciting environments where students explored concepts and developed their mathematical thinking in a wide range of topics. Nearly all teachers used the EM curriculum exclusively and were successful at implementing most of the aspects of the program. While largely positive about the curriculum, many teachers also indicated that it took several years to change the focus of teaching from the old practice-based approach to the new problem-solving approach. Additional research is needed to investigate how to help teachers further implement exemplary *Standards*-based classrooms.

References

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

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

Carroll, W. M. (1998). Geometric knowledge of middle school students in a reform-based mathematics curriculum. School Science and Mathematics, 98 (4), 188–197.

Carroll, W. M., Fuson, K. C., and Diamond, A, (2000). Use of student constructed number stories in a reform-based curriculum. Journal of Mathematical Behaviour, 19:49–62.

Drueck, J. V., Fuson, K. C., Carroll, W. M., and Bell, M. S. (1995, April). Performance of U.S. First Graders in a Reform Math Curriculum Compared to Japanese, Chinese, and Traditionally Taught U.S. Students. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Drueck, J. V., Fuson, K. C, and Carroll, W. M. (1999). Performance of U.S. first Graders in a reform math curriculum. Paper submitted for publication.

Educational Testing Service. (1990). Math background question: 1990 NAEP release items. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (1992). Math background question: 1992 NAEP release items. Princeton, NJ: Educational Testing Service

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

Fraivillig, J.L. (2001). Strategies for advancing children's mathematical thinking. Teaching Children Mathematics, 7 (8): 454–459.

Fuson, K. C., Carroll, W. C., and 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.

Fuson, K. C., Ding, D. and Perry, M. (1997, March). Teacher questions in EM second-grade classes compared to teacher questions in Japanese, Taiwanese, and traditional U.S. classrooms. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Kenney, P. A. & Kouba, V. L. (1997). What do students know about measurement? In P.A. Kenney and E.A. Silver (Eds.), Results from the sixth mathematics assessment of the national assessment of educational progress (pp. 141–164). Reston, VA: National Council of Teachers of Mathematics.

Lindquist, M. M. (Editor) (1989). Results from the fourth mathematics assessment of the national assessment of educational progress. Reston, VA: National Council of Teachers of Mathematics.

Lindquist, M. M. & Kouba, V. L. (1989). 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.

McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., & Cooney, T. J. (1987). The underachieving curriculum: Assessing U.S. school mathematics from an international perspective. Champaign, IL: Stipes.

Mills, V. L., Wolfe, R., and 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, IL.

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

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

National Council of Teachers of Mathematics. (1995). Assessment standards for school mathematics. Reston, VA: The Council.

Okamoto, Y. Muira, 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. Muira, I. T., Tajika, H. and Takeucki,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, D.C.: National Center for Educational Statistics.

Psychological Corporation. (1992). Stanford achievement test, primary 2, form L. Harcourt, Race Jovanovich.

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.

Silver, E.A., Shapiro, L.J., & Deutsch, A. (1993). Sense making and the solution of division problems involving remainders: An examination of middle school students' solution processes and their interpretations of solutions. Journal of Research in Mathematics Education, 24(2), 117–125.

Stevenson, H. W. & Stigler, J. W. (1992). The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education. New York: Simon and Schuster.

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

Stodolsky, S. S. (1988). The subject matters: Classroom activity in math and social studies. Chicago: University of Chicago Press

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

Appendices

Appendix A.1

First-grade study: Percentage correct for Japanese, Chinese, U.S. traditional, and EM first graders

FALL WHOLE CLASS TEST	Japanese	Chinese	U.S. Traditional	EM
1. 5 + 1	98 >	96 =	87 <	93 =
2. Here are 5 birds and 4 sets of bird cages. Circle the set of birdcages that has one birdcage for each bird.	98 >	91 =	84 =	87 =
3. 9 – 1	80 >	74 >	52 =	53 =
4. How many dots are there? Count them and write the number. (17 dots in a row)	96 >	96 >	78 <	89 =
5. 5 + 4	99 >	96 >	77 <	85 =
6. 8 – 3	87 >	71 >	45 <	51 =
7. 6 – 0	91 >	77 =	83 >	76 =
8. Complete pattern: 73, 74, __, __, 77, 78, 79, __	80 =	71 <	41 <	82 =
9. 14 + 5	78 >	68 =	35 <	63 =
10. 14 – 6	35 =	32 =	15 <	35 =
11. How much of the circle is shaded. Finish the fraction 1 /__. (One-fourth shaded).	2 <	11 <	6 <	23 =
SPRING TEST AND INTERVIEW WORD PROBLEMS				
1. Joey had 3 marbles and then found 2 more. How many marbles does Joey have now?	98 =	97 =	89 =	93 =
2. Jane's father gave her 6 cookies. She ate 2 of them. How many did she have left?	93 >	81 =	73 =	81 =
3. Some squirrels picked up 9 nuts yesterday and 4 nuts today. How many nuts do they have altogether?	88 =	76 =	64 <	79 =
4. There were 15 bunnies. 9 hopped away. How many bunnies were left?	66 =	38 <	30 <	56 =

Table continues on next page)

Appendix A.1 continued

	Japanese	Chinese	U.S. Traditional	EM
5. Lisa invited 4 friends to a party. Then she invited 3 more friends. But 2 friends count not come. How many of Lisa's friends came to the party?	68 =	50 =	42 <	60 =
6. You had 6 jellybeans and your friend has 4 and you both want to have the same number. How many would you give to your friend?	58 >	44 =	17 <	32 =
7. One day there were some apples under a tree. The next day, 6 more apples fell to the ground. Now there are 14 apples. How many apples were on the ground the first day?	21 =	13 =	3 <	14 =
8. Mary measured a tree and it was 139 inches tall. A few months later it was 168 inches tall. How much had it grown since Marty measured it the first time?	6 =	2 =	1 <	9 =
9. Kate is 3 years older than her brother. How many years older than her brother will Kate be in 5 years?	23 >	16 =	3 <	14 =
10. I have 50 envelopes and 60 sheets of paper. What is the largest number of letters that I can mail?	46 =	44 =	19 <	39 =
11. Chris has 26 toy cars. Mary has 19. How many do they have in all?	29 =	25 =	13 <	37 =
NUMBER CONCEPTS				
1. Write 57	96 =	98 =	92 <	98 =
2. Write 10 more than 57.	50 <	28 <	31 <	68 =
3. Circle the number that shows 132. (10032, 132, 32)	70 =	38 <	25 <	62 =
4. Circle the biggest number. (12, 102, 120, 21)	91 =	87 <	76 <	94 =
5. Circle the biggest number. (502, 5000, 520, 900)	92 =	73 <	74 <	86 =
6. Draw a circle around half of the stars. (12 stars not in array)	52 >	30 =	11 <	32 =
7. Write the number that is the same as ten tens.	17 <	30 =	10 <	37 =
8. How can you arrange these 3 digits make the biggest number? (3, 6, 1)	40 <	28 <	28 <	63 =

Appendix A.1 continued

EQUATIONS	Japanese	Chinese	U.S. Traditional	EM
1. What sign should go in the space? 6 __ 3 = 3?	87 >	69 =	61 =	67 =
2. Circle the number story that best shows "John had 6 marbles. He gave 2 marbles to Sam. Now he as 4 left." (6 – 2 = 4, 6 – 4 = 2, 4 – 2 = 6)	92 =	58 <	80 <	90 =
3. 4 + 6 = ___ + 4	30 >	9 =	15 =	16 =
4. 10 + ___ = 10	59 =	31 <	34 <	62 =
5. Which of the numbers 0, 2, 3, 5 can fit in both spaces. 10 – __ – ___ = 6	52 =	18 <	15 <	41 =
ESTIMATION				
1. About how many paper clips lined up would be as long as this pencil?	75 =	61 =	51 <	70 =
2. There are three lines. Which shows 2 inches?	51 >	39 =	31 =	35 =
3. (1 and 100 on number line. X around 33.) What number do you think X is (23 – 43 accepted.)	11 =	10 =	4 =	7 =
4. (80 and 100 on number line. X around 85.) What number do you think X is? (83 – 87 accepted.)	66 =	34 <	40 <	69 =
TABLES AND GRAPHS				
1. Comparing numbers in a table.	79 =	54 <	55 <	70 =
2. Interpreting a table 1.	81 >	53 =	81 =	85 =
3. Interpreting a table 2	96 >	90 =	81 =	85 =
3. Interpreting a table 3	84 >	69 >	42 =	48 =
4. Reading a bar graph	40 =	26 =	23 <	38 =

Notes: The >, =, < compare performance of the named sample to the EM sample, $p < .01$. For example, > means that this sample did significantly better than the EM sample on that question. For exact questions, see Drueck, Fuson, and Carroll, 1999 and Stigler, Lee, and Stevenson, 1992.

Appendix A.2
NAEP comparison—Percent correct on each item and subtest

	4^{th} grade EM	4^{th} grade NAEP sample
NUMBERS AND OPERATIONS		
Symbolic computation		
1A. 503 – 207	65	53
2B. 604 – 207	56	62
Mean on computation subtest	62	58
Story problems		
3A. Carol buys a ball for 55 cents and a game for 37 cents. How much change should she get back from $1?	79	45 *
4A. Chen has $10 to buy a model plane, glue, and paint as shown (3 pictures with prices given.) At which of the following times could an estimate have been used instead of exact numbers?	67	45 *
5A. Mrs. Garcia bought 5 dozen eggs a $.89 a dozen. What is the total cost of the eggs?	82	56 *
6A. If 1 1/3 cups of flour are needed for a batch of cookies, how many cups of flour will be needed for 3 batches?	50	21 *
7B. Jill needs to earn $45 for a class trip. She earns $2 each day on Mondays, Tuesdays, and Wednesdays, and $3 each day on Thursdays, Fridays, and Saturdays. How many weeks will it take her?	59	22 *
8B. Carol wanted to estimate the distance from A to D along the path shown on the map below. She correctly rounded each of the given distances to the nearest mile and then added them. Which of the following could be hers?	48	25 *
9B. Christy has 88 photographs to put in her album. If 9 fit on each page, how many pages will she need?	56	37 *
Mean on story problem subtest	65	36
Place value questions		
10A. How much would 217 be increased if the digit 1 were replaced by the digit 5?	69	36 *
11A. Which of the following is closest to 15 seconds? (14.1, 14.7, 14.9, 15.2)	91	63 *

Appendix A.2 continues on next page

Appendix A.2 continued

	4th grade EM	4th grade NAEP sample
12B. The census showed that three hundred, fifty-six thousand, ninety-seven people lived in Middleton. As a number this is?	94	72 *
13B. By how much would the value of 5,647 be decreased of the 5 were replaced by a 2?	89	61 *
14B. What number is four hundred five and three-tenths	91	69 *
15B. Which of the following represents "nine-tens?"	70	67
Mean on place value questions	84	61
Mean on Numbers and operations subtest	72	49
ALGEBRA AND PATTERNS		
16A. Children's pictures are hung in a line as shown in the figure above. How many tacks are needed to hang 28 pictures?	36	25
17A. If ☐ represents the number of newspapers that Lee delivers each day, which of the following represents the total number of newspapers that Lee delivers in 5 days? (5 × ☐)	79	48 *
18B. Inducing a non-linear pattern from a table (Puppy's age and weight)	78	32 *
Mean on algebra subtest	62	35
GEOMETRY		
19A. Which letter has 2 parallel lines? (A, T, K, and N)	61	27 *
20A. In the space below, use your template to draw a square with two of its corners at the points below.	91	40 *
21A. In the space below, draw a rectangle 2 inches wide and 3 ½ inches long.	67	18 *
22B. Which streets appear to be parallel (diagram)	86	49 *
Mean on geometry subtest	75	34 *
MEASUREMENT		
23A. A rectangular carpet is 9 feet long and 6 feet wide. What is the area of the carpet in square feet?	59	19 *

Appendix A.2 continued

	4th grade EM	4th grade NAEP sample
24A. The weights of 3 objects were compared using a pan balance. Two comparisons were made as shown in the figure above. Which object is the heaviest?	75	42 *
25B. What is the area of the rectangle (4 * 6 with dimensions given and square units shown)	80	31 *
26B. Henry is older than Bill, and Bill is older than Peter. Which statement is true? (Comparison of ages)	69	62
27B. On the grid below, draw a rectangle with an area of 12 square units	80	43 *
Mean on measurement subtest	71	39
DATA ANALYSIS (Statistics)		
28A. Reading a bar graph (closest to 400)	80	86
29B. In a bag of marbles, ½ are red, ¼ re blue, 1/6 are green, and 1/12 are yellow. If a marble is picked, which color is most likely?	64	25 *
30B. Reading a bar graph	48	49
Mean on analysis and graphs	67	53
Mean on All NAEP 29 Items	71	44

Note: A and B denote the Form of the test taken by EM students. For Form A, n = 106; for Form B, n = 64. Because of the different number of EM students taking the two tests, EM subtest scores are weighted averages. In the NAEP sample, approximately 2700 students answered each question on the 1990 test and 2000 answered each on the 1994 test.
* Denotes a significant difference on the Chi-square test, $p < .01$.

Appendix A.3

Percentage Correct for Japanese, Chinese, U.S. Traditional and U.S. Reform (EM) Children: Fifth grade results

Items	Japan-ese	Chi-nese	U.S. Trad	U.S. EM
Word Problems				
Kim's weight is 49.7 pounds, Sue's weight is 50.4 pounds. Who is heavier?	100	100	94	98
How much?	74	70	26 <	64
8 children went on a picnic. Each child took 2 sandwiches and 3 cookies. Altogether how many cookies did the children take?	98	85	75 <	92
33 people went to a football game. They went home in 7 cars. In 6 of the cars, there were 5 people each. How many people were in the 7th car?	92	82	54 <	84
Dad cut a cake into 16 pieces. George ate one-fourth of them. How many pieces were left?	65	63	30 <	78
A truck will hold 33 boxes of oranges. How many trips will be needed to carry 152 boxes of oranges to a store?	52	54	15 <	58
A field was 20 meters long and 15 meters wide. How long is the fence that goes completely around the field?	42	18	15 <	33
A train left Chicago at 1:55 p.m. and Arrived at St. Louis 6 hours and 52 minutes later? What time did it arrive in St. Louis?	34	27	3 <	42

Appendix A.3 continues

Appendix A.3 continued

Percentage Correct for Japanese, Chinese, U.S. Traditional and U.S. Reform (EM) Children: Fifth grade results

Items	Japanese	Chinese	U.S. Trad	U.S. EM
Number concepts and equations				
Solve the problem				
4 + 6 + 3 = ___ + 3	90 >	85 >	39 <	70
Here are 5 digits, 2, 6, 3, 5, 1. How could you arrange them to form the smallest number? (Use each digit only once.)	93	77	57 <	86
Draw a circle around one-half the stars. (12 stars not in an array.)	93	83	71	83
What sign goes in the box to make this number sentence true? 5 □ 4 = 3 × 3	92	86	64 <	90
Put parentheses to show how this problem was solved. 48 = 4 × 2 + 4 × 2	64	53	30 <	69
Read the numbers:				
30%	100	81 <	92	100
800070	96	90	53 <	94
20350076	96	76 <	45 <	96
Write three-fourths	100	95	93	100
Can you tell me another way to say ½?	94	75	40 <	88

Appendix A.3 continues

Appendix A.3 continued

Percentage Correct for Japanese, Chinese, U.S. Traditional and U.S. Reform (EM) Children: Fifth grade results

Items	Japanese	Chinese	U.S. Trad	U.S. EM
Operations and mental computation				
What is 3 × 6?	100	100	98	100
What is 30 × 60?	73	74	35 <	76
What is 300 × 600?	76	65	13 <	63
(all problems below in vertical format)				
15 + 26	90	96	82	94
415 + 376	88	92	75 >	76
28 + 73 + 54	75	70	63	78
409 – 156	74	71	53 <	82
28 × 14	44 >	44 >	9	18
Geometry				
One of these is a right angle. Circle it. (4 angles shown)	93	97	54 <	97
Find the area. (Rectangle with area of 24 sq. cm).	91 >	91 >	33 <	77
Circle the trapezoid. (4 figures shown.)	96	92	27 <	95
Find the area. (Rectangle Labeled 24 ft by 8 ft.)	88 >	88 >	8 <	69

Appendix A.3 continues

Appendix A.3 continued

Percentage Correct for Japanese, Chinese, U.S. Traditional and U.S. Reform (EM) Children: Fifth grade results

Items	Japan-ese	Chi-nese	U.S. Trad	U.S. EM	
Find the volume. (Rectangular prism with sides 4m, 2m, 5m.)		97	88	5	33
	>	>	<		
Find the measure of angle A. (Triangle with interior angles labeled 50° and 45° given)	40	74	0	38	
		>	<		
Find the sum of angles A, B, and C. (Interior angles of triangle)	53	73	1	34	
	>	>	<		

Notes: The > means that the score of that sample was significantly better that that of the EM sample; the < means that the score of that sample was significantly lower than that of the EM sample (p < .01, Chi-square test).