Submission to the Program Effectiveness Panel
U.S. Department of Education

Comprehensive School Mathematics Program

ABSTRACT

Goals. The primary goal of the Comprehensive School Mathematics Program is to provide a complete K-6 Mathematics program for students of all ability levels which develops a broad and balanced range of skills. Students will be actively involved in the world of mathematics not simply drilled in the techniques of arithmetic. They will understand the content and applications, develop techniques for learning them, and use their mathematics to solve problems.

Needs. The needs addressed by this program have been discussed recently in reports by many national groups, including the National Council of Teachers of Mathematics, the National Science Board, and the National Assessment of Educational Progress. These reports consistently stress several things:

- Problem solving should be the focus of school mathematics.
- The study of mathematics should emphasize developing higher order thinking skills (reasoning, analyzing, estimating, inferring, etc.), understanding of concepts, communicating about mathematics, making mathematical connections, and applying mathematics.
- Basic skills in mathematics should be defined to include more than computational facility.
- School mathematics should provide for an integrated study with increased emphasis on content such as geometry, measurement, patterns, relations, numeration, probability, statistics, logic, algorithmic thinking, and applications.
- Mathematics programs should take advantage of calculators and computers.

The CSMP curriculum is responsive to these and other concerns and provides an alternative to present mathematics curricula.

Method of Operation. CSMP is designed to be used by the regular classroom teachers of grades K-6. It does not require additional personnel or facilities, beyond the need for a CSMP coordinator, usually a district mathematics coordinator, to manage the program. Teacher preparation is recommended and several sources of training and implementation assistance are provided, including a Coordinator's Manual for districts wishing to conduct their own inservice. Materials include very extensive teacher's guides with detailed plans for all lessons, storybooks, student worksheets and workbooks, manipulatives and tools, and other demonstration materials. The schedule is organized in a spiral fashion. CSMP makes use of a number of special pedagogical devices - non-verbal "languages" - to aid in understanding the content and methods of problem solving. These include the Papy Minicomputer, string pictures, and arrow diagrams. The content includes significant attention to probability and statistics, geometry and measurement, numeration and number sense, and logical thinking.

Audience. The intended audience of CSMP is all K-6 students; that is, CSMP is designed to be used as the mathematics curriculum by students at all ability levels in grades K-6.

Claims. Two claims are made in this submission:

1. CSMP improves students' abilities to use the mathematics they have learned in new problem situations involving estimation, mental arithmetic, representations of numbers, number patterns and relationships, word problems, and producing multiple answers.

2. CSMP students perform in traditional computation skills as well as comparable non-CSMP students.

BASIC INFORMATION

A. Project Title, Location, Contact Person

Project Title: Comprehensive School Mathematics Program (CSMP)

Applicant Agency: Mid-continent Regional Educational Laboratory
2550 S. Parker Road, Suite 500
Aurora, CO 80014

Contact Person: Clare Heidema, (303) 337-0990

B. Original Developer/Applicant Agency

Developed by CEMREL, Inc., a non-profit educational laboratory in St. Louis, Missouri, with completion of revision stages, updates, and enhancements made by McREL.

McREL is a non-profit educational laboratory with offices in Aurora, Colorado and Kansas City, Missouri. Its mission is to create a community of interest among those individuals and organizations interested in learning how to help schools meet the needs of an everchanging society. The key strategy the Laboratory uses is to identify and encourage the use of knowledge that improves education. Laboratory activities include dissemination of information, product development, networking, training, and technical assistance.

Key People: Burt Kaufman, Director, CSMP, through 1979
Clare Heidema, Director, CSMP, from 1979
Martin Herbert, Director, Evaluation Studies

C. Years of Project

Dates developed and evaluated: 1972-1984
Dates operated: 1972-present
Dates disseminated: 1979-present

D. Sources of Funding

Development and evaluation funds were from the National Institute of Education/Office of Educational Research and Improvement. Dissemination funds were from the National Diffusion Network.

Development and Evaluation 1972-1984: Approximately 8½ million dollars

Dissemination 1979–1984: Approximately $200,000
1984–1988: $235, 645
1988–1990: $114,963
1990–1992: $162,733

DESCRIPTION OF THE PROGRAM

A. Goals

a. To provide a complete and up-to-date mathematics curriculum, K-6, for students of all abilities.

b. To draw children's attention to the world of mathematics and not exclusively to drill them in the techniques of arithmetic.

c. To develop a balanced approach to concepts, skills, and applications.

d. To improve teaching through inservice preparation and in-depth teacher manuals.

f. To encourage a three level approach to learning mathematics: understanding the content and applications, developing the techniques and processes for learning that content, and applying the appropriate means in the solution of problems.

g. To improve students' abilities to apply their mathematics to unfamiliar mathematical situations.

B. Purposes and Needs Addresses

A number of national bodies have made extensive recommendations for mathematics education:

- The National Council of Teachers of Mathematics
  (Curriculum and Evaluation Standards for School Mathematics, 1989; Professional Standards for Teaching Mathematics, 1991)

- The National Assessment of Educational Progress
  (Fourth and fifth Mathematics Assessments, 1986 and 1990; The State of Mathematics Achievement, 1991)

- National Research Council: Mathematical Sciences Education Board
  (Everybody Counts: A Report to the Nation on the Future of Mathematics Education, 1989; Reshaping School Mathematics: A Philosophy and Framework for Curriculum, 1990)

- Educational Testing Service
  (The Mathematics Report Card: Are We Measuring Up? 1988)

- American Association for the Advancement of Science
  (Project 2061: Science for All Americans, 1989)

The recommendations of these organizations may be summarized as follows:

- Problem solving should be the focus of school mathematics.
- Basic skills in mathematics should be defined to include more than computation.
- The study of mathematics should emphasize reasoning and thinking skills, developing and understanding of concepts, communicating about mathematics, and applying mathematics.
- School mathematics should provide for an integrated study with more emphasis on numeration and number sense, patterns and relations, geometry, measurement, probability, statistics, algebraic and algorithmic thinking, and mathematical connections.
- Mathematics programs should take advantage of calculators and computers.

CSMP offers a response to these concerns and a unique alternative to present mathematics curricula. For the features listed below, it was profiled by the National Commission on Excellence in Education.

- a complete K-6 curriculum for all students
- a problem solving orientation
- a program of expanded basic skills including higher-order thinking skills

- a unified study of mathematics, not just arithmetic
- well-tested teaching methods with extensive teacher inservice
- situational and discovery learning as a spiral approach
- research and classroom based with five years of development and evaluation at each grade level

C. Intended Audience

CSMP is designed to be used as the mathematics curriculum by students of all ability levels in grades K-6.

D. Background, Foundation, and Educational Framework

Comprehensive School Mathematics Program stands for both the name of a curriculum, CSMP, and the name of a project, responsible for the development of curriculum materials. The Project was established in 1966 under affiliation with Southern Illinois University, Carbondale, IL, and was later incorporated into CEMREL, one of the national educational laboratories funded at the time by the U.S. Office of Education. The portion of the project devoted to the elementary curriculum began formally in 1970. The development was heavily influenced by the work of Mme Frederique Papy, who joined CSMP in 1972 to direct the research and development effort.

At each grade level, a five-year research and classroom based development/evaluation/revision cycle was followed, on a staggered basis.

> Year 1. Instructional materials were planned and taught by CSMP staff to heterogeneous public and parochial school classes. This experience led to a Local Pilot Test version of the materials.
>
> Year 2. The Local Pilot Test materials were used by about ten regular classroom teachers in St. Louis area schools. CSMP staff observed the classes and revised the materials, producing an Extended Pilot Trial version.
>
> Years 3 and 4. The Extended Pilot Trial version was used for two years in a national network of cooperating schools. Extensive evaluation data, including comparisons of CSMP and non-CSMP classes, was collected.
>
> Year 5. Revisions based on Local and Extended Pilot Test data resulted in the versions of the materials which were then readied for publication.

An extensive evaluation dealing with many aspects of CSMP usage was conducted by an independent unit within CEMREL, drawing on the expertise of an external group of nationally recognized evaluation consultants. This work led to the publication of many formal evaluation reports (about 60 volumes).

In 1978, CSMP's K-3 curriculum was approved by the Joint Dissemination Review Panel as a nationally validated program, marking the beginning of full scale dissemination and implementation. The full K-6 program was approved in 1984.

Among the basic principles which guided developers were the following:

- Mathematics is a unified body of knowledge and should be organized and taught as such.

- Mathematics as a body of knowledge requires certain ways of thinking and cannot be done by the exclusive use of memory.

- Children learn through interrelated experiences and by reacting to problem situations.

- Mathematics is best learned when applications are presented which are appropriate to students' levels of understanding and to their natural interests.

One of the manifestations of these convictions in the construction of the CSMP K-6 curriculum is the spiral approach. The content is completely sequenced in spiral form so that a student is brought into contact with each area of content continuously throughout the program. This approach consciously precludes atomizing the content and mastering each bit before continuing to the next. Students work through repeated exposures to the content, building interlocking experiences of increasing sophistication.

The content is learned in an atmosphere of constant connections with applications, from simple story situations to challenging applications to nontrivial simulations of real world problems. The emphasis is at all times on a two-level approach to learning: understanding the content itself and its applications, and equally important, developing the techniques and processes of learning the content. It is the latter form of knowledge that gives power to apply the former.

To this end, the content is presented as an extension of experiences children have encountered in their development, both at the real-life and fantasy levels. Using a "pedagogy of situations" students are led through sequences of problem-solving experiences presented in game-like and story settings. Powerful nonverbal "languages" allow the students immediate access to the mathematical ideas and methods necessary not only for solving these problems but also for continually expanding their understanding of the mathematical concepts themselves. These languages include:

> The language of strings: this "language" mimics a basic method of organizing and collecting data and deals with the fundamentally useful and important mathematical notion of set.
>
> The language of arrows: this graphical language models the fundamental processes involved in comparing and analyzing sets and operations on them. That is, it deals with relations and functions including, but not restricted to, permutations, ordering relations, and the numerical processes of adding to, subtracting from, multiplying by, and sharing equally among.
>
> The language of the Papy Minicomputer: the Minicomputer is an extremely effective abacus invented by Georges Papy, a noted Belgian mathematician, that models the positional structure of our system of numeration and lends itself to very sophisticated algorithmic processes, including the standard algorithms for the basic numerical operations.

Other tools and manipulatives such as the calculator, various geometry tools, random devices, various kinds of blocks, counters, tiles, etc. are used extensively throughout the curriculum to pose problems, explore concepts, develop skills and define new ideas.

Another tenet of CSMP's development philosophy is that no single method of classroom management can meet the needs of every student. Hence the program was constructed to allow numerous opportunities for whole class participation, small group participation, and independent individual study.

CSMP was designed to teach students mathematics and not merely arithmetic. One of the key aspects of CSMP has been its dual emphasis on both mathematical content and pedagogy designed to support mathematical reasoning. As the program was developed, piloted, and revised, both content and pedagogy were modified to reflect classroom experiences.

E. Features: How the Program Operates

1. Scope

CSMP is a complete mathematics program for all students. It has also been successfully adapted to many specialized audiences such as gifted students, compensatory education groups, and bilingual populations. It is also used as components of programs for special populations at all ability levels.

2. Curriculum

The curriculum is divided into four levels: Kindergarten; Grade One; Upper Primary Grades (grades 2-3); and Intermediate Grades (grades 4-6). In kindergarten and first grade the content is organized and presented in a single sequence of lessons emphasizing elementary numerical and geometric concepts and their exemplification in the CSMP languages and tools. In the other grades the content is organized by four interrelated strands:

- The World of Numbers
- Geometry and Measurement
- Probability and Statistics
- The Language of Strings and Arrows (logical thinking, reasoning, and related numerical skills)

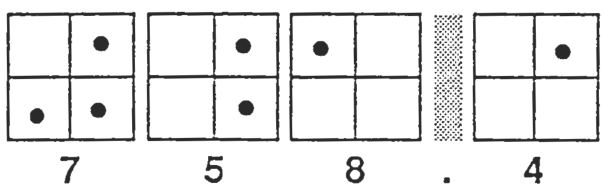
The schedule is organized in a spiral fashion by days of the week. On two days, lessons come from the World of Numbers strand. Two days are devoted to the other strands, the emphasis varying by grade level. One day is devoted to special workbooks which provide practice and problems from recent lessons. Each grade level has its own prescribed schedule of lessons which is presented in the Teacher's Guide.

The Teacher's Guides present a series of very detailed lessons, which describe the materials needed, pictures to be drawn, questions to ask students, and handling of student responses. A typical lesson will consist of a group lesson, devoted to a class discussion of a particular mathematical situation, followed by individual or group work, often with prescribed worksheets. The worksheets are written at increasing levels of difficulty; some students will complete more worksheet pages than others, allowing for individualization by the teacher and the inclusion of high level material. The teacher's role in the group lesson is to present the situation and pose questions that will guide students to a desired mathematical experience or conclusion.

3. Learner Activities

This section will present a brief flavor of some of the CSMP activities. It should be reemphasized that these ideas are presented in a unified and integrated manner using the spiral approach. (Unfortunately, the use of color, prevalent in CSMP activities, is not possible in this black and white submission.)

The World of Numbers. The Minicomputer, a simple abacus consisting of colored squares on which checkers are placed, combines the usual positional notation with the fundamental notion of doubling. It is suitable for the introduction of concepts of place value, adding, subtracting, multiplying and dividing by "small" numbers, as well as decimal numbers and negative numbers.



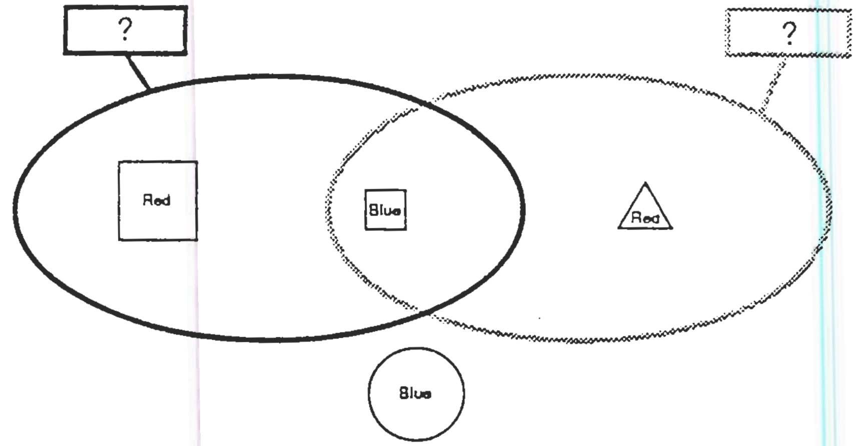
Another tool used extensively and creatively in the curriculum is the calculator. It is used by students for several purposes: to complement computation; develop process skills and understanding; explore patterns and decimal concepts; pose problems and new ideas; provide opportunities for discovery learning; make and test conjectures; and develop estimation and mental arithmetic skills.

Geometry and Measurement. These lessons comprise about 20% of the curriculum with early emphasis on measurement. The metric system is used extensively but not exclusively. There are activities which involve properties of various shapes, parallelism, topological ideas of inside and outside, transformations and symmetry, spatial perception, and other geometric ideas. One interesting innovation is the use of taxi-geometry, studied on a a square grid, which introduces the notion of taxi-distance as the shortest distance following the lines on the grid. Taxi-distances often differ from Euclidean ("as the crow flies") distances and thus serve to highlight the special properties of our usual intuitive geometry.

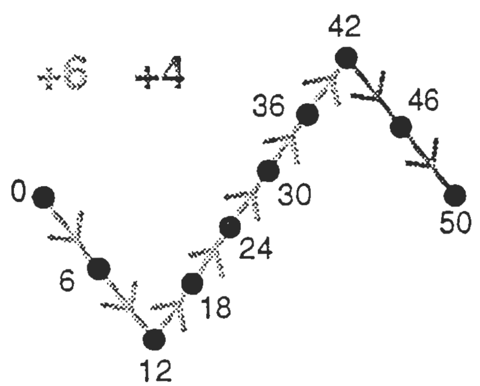
Probability and Statistics. Throughout the program there are many game, story, and problem situations which involve the students with basic ideas of probability and statistics. Some involve data collection with a view to statistical analysis. Others involve prediction and chance events where probabilities play a fundamental role. These ideas are unified with other content areas by making use of geometric methods, by invoking number patterns, and by providing support to work with negative numbers and fractions.

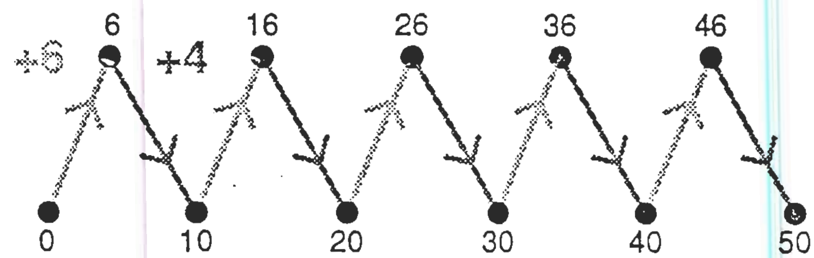
Classification. Colored "strings" provide a powerful non-verbal language for classification which can yield immediate access to significant logical thinking exercises. For example, given the objects already classified below, what is the correct label for the dark and for the light strings?

The Dark string is one of these:	The Light string is one of these:
Red	Red
□	□
△	△
Blue	Blue
○	○
Big	Big
Small	Small



Relations and Functions. Colored arrows are used to represent mathematical relationships, providing a precise, easily understood, non-verbal way of working with relations. One way this "language" is used is to help students visualize strategies in problem-solving situations. Below are two solutions to the problem, "Pencils cost 6 cents each and erasers 4 cents. How can we spend exactly 50 cents on pencils and erasers?"





We could buy 7 pencils and 2 erasers.

We could buy 5 pencils and 5 erasers.

4. Learning Materials

Included in the curriculum package at each grade level are teacher's guides, worksheet booklets, storybooks, workbooks, and assorted manipulative and demonstration materials. To illustrate the extent of the materials, at fourth grade there are ten 32-page student workbooks, 330 pages of student worksheets, six storybooks, and four volumes of teacher's guides averaging 400 pages in length. A Coordinator's Manual and individual training kits for teachers are also available; these materials allow teachers to experience selected lessons in much the same manner as their students do, and provide extensive guidance on questions of classroom management, organization, testing, etc.

5. Staff Development Activities

CSMP offers inservice workshops (graduate credit may be arranged) for all teachers involved in program implementation in one of several ways: one week programs for district coordinators who will in turn train their local teachers; direct inservice workshops at district locations; and courses at some educational institutions based on the CSMP inservice workshop model.

6. Monitoring and Evaluation Activities

A continuous monitoring process allows the recording of involvement in activities and the assessment of progress. Workbooks, containing problems of varying difficulty, provide an instrument which can be used to assess the progress of students on a week-by-week basis. Observation and progress charts are provided to help teachers keep track of student performance and progress.

F Significance of Program Design

CSMP has many unique features that distinguish it from other mathematics curricula. Its design offers an alternative that is a complete curriculum - not just a management system, not a collection of supplementary activities, not dependent on commercial text materials, not just for special populations. Many programs that aim at improving problem-solving skills, developing higher-level cognitive skills, incorporating concept tools, making mathematical connections, etc., are designed as an add on to a basal text. Several needs statements in the introduction to NCTM's Curriculum and Evaluation Standards for School Mathematics apply as well to CSMP:

> "The driving force for the development of (the) standards is a vision of the mathematics all students should have an opportunity to learn and the way in which instruction should occur. Classrooms should be places where interesting problems are explored using important mathematical ideas... This vision sees classrooms in which students are actively engaged in making conjectures and discussing ideas... Finally, this vision sees teachers encouraging students and probing for ideas."

POTENTIAL FOR REPLICATION

A. Settings and Participants

Materials were developed in intact classrooms in the Carbondale, Illinois, and University City, Missouri, school districts. Both were racially integrated (20-50% black), middle class communities and both, because of different circumstances, had disproportionately high numbers of both high and low ability students.

Local Pilot Testing (approximately ten classes at each grade level) occurred most frequently in five St. Louis area school districts, including the City of St. Louis, one "inner" suburban, two "outer" suburban, and one "exurban" district. Two of these districts had over 50% black enrollment (both in the school district as a whole and in most of the pilot classes) and could be classified as lower or lower-middle class communities. At the other extreme, one district was primarily white with a high SES. These local pilot tests involved regular classroom teachers in intact classes of students. Selection of participating classes was done jointly by CEMREL and the school districts, usually on a school basis (all teachers at a given grade level in the school participated). In higher grades, teachers more or less inherited their students and CSMP from the previous year.

Extended Pilot Testing, the vehicle which provided the most definitive evaluation data about CSMP usage and learning, was conducted on a national basis. Joint agreements with school districts allowed for, among other things, the collection of extensive data comparing the learning of CSMP and non-CSMP students. A total of 27 school districts participated in these comparison studies, at least nine districts per grade level with some districts participating at more than one grade. These 27 districts were distributed as follows:

Type of Community		Geographic Location	
7	large city	7	east
12	suburban	8	central
4	medium city	6	upper midwest
4	small city/rural	3	south
		3	west

Altogether, about 150 CSMP classes participated. The comments above in the description of Local Pilot Testing, regarding teacher/class selection and inheritance of classes, also apply to the Extended Pilot Tests.

The intent of the testing was to thoroughly evaluate the materials under realistic conditions with a wide spectrum of students, teachers, schools, and communities, and this objective was accomplished.

B. Replicable Components and Documentation

All materials, including student, teacher, and coordinator materials, are readily available to prospective CSMP users, as are a wide variety of ancillary materials relating to training, special-purpose usage of materials, supplementary materials, and so forth. These materials are fully described in several sources. A network of turnkey trainers and cooperating educational agencies exists to help school districts in their preparation for and use of the curriculum. McREL itself, supported by the National Diffusion Network, is the primary vehicle for dissemination and maintains a hot-line and staff devoted to CSMP.

C. User Requirements

The program is to be taught by the regular classroom teacher. No other personnel are required, nor is any equipment or facility beyond the normal classroom. To adopt CSMP, a school district must sign an adoption agreement, appoint a CSMP coordinator (normally the district mathematics supervisor), and agree on an implementation plan that provides for the training of teachers, the evaluation of the program, technical assistance and support services. Depending on the grade level of implementation, between 6 and 30 hours of training are recommended.

D. Costs

Materials costs are shown below for three grade levels. Start-up costs refer to non-consumable materials, such as teacher and coordinator materials and certain of the classroom and student materials. The useful life of these materials will vary widely, but might be estimated to be five years. Operational costs refer to consumable items that need to repurchased each year. All figures are based on a per student computation.

Grade	Start-Up	Operation
K	$4.40	$1.40
3	$5.30	$7.20
6	$5.30	$8.00

There are no special equipment costs. Personnel costs are difficult to quantify, because of the variety of options available for teacher training. There are essentially two components - teacher training and program management. All teacher training can be done by the local coordinator, using days or half days that are part of the district professional development program, thus incurring only small costs. At the other extreme, teachers can be brought into a full week summer training session conducted by an outside CSMP consultant, an option which costs considerably more. Program management - teacher support, CSMP liaison, materials ordering, evaluation, and representing the program to the district school board, administration, and parents - is the responsibility of the CSMP coordinator, usually a district mathematics coordinator. Hence these activities are usually a normal part of that persons job.

EVIDENCE

A. Claims

1. CSMP students are better able than comparable non-CSMP students to apply the mathematics they have learned to new problem situations using processes involving:

   Estimation
   Mental Arithmetic
   Representations of Numbers
   Number Patterns and Relationships
   Word Problems
   Producing Multiple Answers

2. CSMP students perform in traditional computation skills as well as comparable non-CSMP students.

B. Methodology

1. Design

CSMP classes used CSMP for their mathematics curriculum during the year of testing. Non-CSMP Classes used the regular district mathematics curriculum, usually one of the major textbook series. Thus the natural unit of analysis for all comparisons was at the classroom level.

Testing was always conducted near the end of the school year, usually in May. All students in the class were tested. Test materials contained three components: computation tests, tests of applying mathematics to new problem situations, and a standardized Vocabulary test (subsequently used as a covariate).

Two designs were used:

Same Teachers - Consecutive Years. In the first year, teachers taught the regular district curriculum and their students were tested at the end of the year. In the following year, the same teachers taught the CSMP curriculum to a new group of students who were tested at the end of the year. The performance of the second group (CSMP) was then compared with the performance of the first group (non-CSMP). This design controlled for effects due to teacher differences.

Different Teachers - Same Year. The school district attempted to match a group of classes studying the CSMP curriculum with a group of classes studying the regular district curriculum. Both groups of classes were tested at the same time near the end of the school year. In some cases all classes in the district were tested, while in other case only a sample was tested. This design did not control for teacher effects but the districts' own interest in obtaining useful comparative data were best served by selecting comparable classes and teachers.

Assignment of students was not random; student differences were statistically controlled through analysis of covariance (see 5. Data Analysis). CSMP and non-CSMP Vocabulary scores are reported in Table 2.

2. Sample

Table 1 lists all districts which conducted comparison studies during 1984-91. The table also shows the type of community, the type of design, whether all classes in the district or only a sample were tested, and the present status of CSMP within the district compared to its status at the time of testing. At each grade level, a comparison study was conducted by at least five districts.

Table 1
Districts Conducting Comparison Studies

District	Type of Community	Design	Classes Tested CSMP/non-CSMP	CSMP Status Since Testing[1]	Grades Tested
Albermarle, VA	County	Different teachers	Sample/Sample	Same	2 3 4 5
Ann Arbor, MI	Small City	Different teachers	All/All	Reduced	3 4 5 6
East Lansing, MI	Small City	Different teachers	All/All	Expanded	2
Guilderland, NY	Suburb	Same teachers	All/Sample	District	5 6
Livonia, MI	Suburb	Different teachers	Sample/Sample	Same	3 5
Manhasset, NY	Suburb	Same teachers	All/All	District	5 6
Ft. Collins, CO	Small City	Different teachers	Sample/Sample	District	2 3 4 5 6
Rockford, IL	Small City	Different teachers	Sample/Sample	Same	2 3 4
St. Joseph, MI	Small City	Same teachers	All/All	District	4 6
Hillsborough, NJ	County	Different teachers	Sample/Sample	Expanded	2

[1] Same: CSMP now at approximately the same level of use as at time of testing
Expanded: CSMP now used in more classes or at more grade levels
District: CSMP now adopted district-wide, grades K-6

Table 2 shows the total number of CSMP and non-CSMP classes tested at each grade level, and the average raw scores for these classes on the Gates-MacGinitie Vocabulary test, administered as part of the testing process. Also shown are the corresponding percentile ranks for these averages. Except at third grade, the difference between the two groups was about one raw score unit out of 45 test items, with CSMP classes usually higher. At third grade the CSMP average was more than two raw score units and this difference was significant at the .04 level.

Table 2

Number Of Classes and Average Vocabulary Scores

Grade	Number of Classes		Mean Score:Vocabulary		Corresponding Percentile Rank	
	CSMP	non-CSMP	CSMP	non-CSMP	CSMP	non-CSMP
2	41	25	33.8	34.7	61	64
3	50	24	35.4	33.1	61	52
4	36	34	29.1	27.9	72	68
5	46	44	32.4	31.4	63	59
6	43	46	27.3	26.2	76	73

3. Instruments and Procedures

Introduction. The tests used in these comparison studies were the MANS Tests (Mathematics Applied to Novel Situations). The tests were developed from 1976 to 1982 at CEMREL by an independent evaluation group. They comprise a series of short tests designed to assess the application of important mathematical processes to problem situations. The tests were developed because suitable standardized mathematics tests for measuring such skills were not available. The need for tests of this kind has been recommended by national organizations such as the National Assessment of Educational Progress (NAEP) in its 1983 report:

> "The very things that are difficult to teach are often difficult or expensive to test. Educational leaders need to pressure test developers to include items that reflect the higher level objectives of the curriculum."

and the National Council of Teachers of Mathematics in its 1989 Evaluation Standards:

> "New tests must be developed to assess problem solving, reasoning, and so on... Without changes in how mathematics is assessed, the vision of the mathematics curriculum described in the standards will not be implemented in classrooms."

Overview. The MANS Tests use standard terminology to present mathematical situations which are relatively unfamiliar to students and do not contain any of the specific language or typical problem activities found in CSMP. Each individual test has its own standardized directions which a specially trained tester uses in explaining the task and sample items to the class. Liberal time limits allow almost all students to finish. For most tests, students produce their own answers instead of selecting from specified alternatives. Any reading requirements are kept intentionally low relative to grade level.

Development. During test development, all tests were reviewed and approved by the external CSMP Evaluation Panel which included distinguished scholars in mathematics, assessment and evaluation, and mathematics education. At each grade level there were two years of test development activity; each year included outside reviews, administration, analysis, and revision. The tests have been used by 46 school districts in 26 states (over 2,000 classes) for both student assessment and curriculum evaluation (including curricula other than CSMP).

Content. Individual MANS tests are grouped together according to the mathematical process entailed by the problem situation in the test. The six mathematical processes correspond to those listed in Claim 2. Appendix A presents sample items for each of these processes or MANS "Categories." Each test is preceded by standardized directions and sample items explained by the tester. At each grade level, the tests are contained in two 16-page booklets. In addition to these special MANS Categories, there was also a Computation category composed of representative items developed from an analysis of items found in the computation tests of five major standardized text series (CAT, CTBS, Metropolitan, Stanford, and Iowa). The MANS tests contain somewhat fewer computation items (average across grade levels equals 27 for MANS versus 38 for standardized tests) and many more concept, problem, or application-oriented items (183 versus 49).

Reliability. Across the five grade levels, there are a total of 85 individual MANS tests. The reliability/ internal consistency (KR20), corrected by Spearman-Brown for an equivalent 20-item test, was above .80 for 72 of the 85 tests; between .75 and .80 for 10 tests; and below .75 for 3 of the tests with a median of .86. Correcting for an equivalent 30-item test, a more usual number for standardized tests, 83 of the 85 tests had a reliability above .80.

Validity. Various correlation coefficients between Total MANS scores and other measures were derived in cooperation with individual school districts at several grades. The median coefficient with standardized reading scores was between .54 and .61 at every grade level, with standardized mathematics scores was .63 (median of all grades), and with teacher estimate of problem solving ability was .59 (measured at 4th grade only). Average teacher rating on a 5-point scale of importance of the various MANS tests (with 5 being "most important") was 4.3 and 4.1 (measured at 4th and 5th grades respectively). Average student rating on a 4-point scale of how well they liked the individual MANS tests was 3.0 (measured at 4th grade only).

Samples. A page of sample items (much abbreviated) is given in Appendix A.

4. Data Collection

Each district appointed a test coordinator to be responsible for all testing activities, including the selection and training of testers who were usually drawn from among the more capable substitute teachers regularly employed by the district. Each tester followed a General Instructions Manual applicable to all grade levels and a Specific Directions Manual for each grade tested. These materials, along with the Coordinator Test Manual, formed the basis for training by the coordinator and for the actual administration of the tests.

At each grade level, the tests were administered in two sessions about a week apart. The sessions lasted from about 35 minutes (second grade) to about 50 minutes (sixth grade). Students wrote their answers directly in the test booklet beside the question. Each page of the 16-page booklets was devoted to a different problem situation.

The tests were then mailed to the central scoring site. A standardized scoring format was used. Trained scorers, as they reviewed each page, entered responses into a data file using a specially developed software program. Samples of data entered were reviewed for adherence to scoring guidelines. All distribution of materials, collection of booklets, supervision of scoring and data entry, and reporting to school districts was done by the same individual who was in charge of the MANS development and the overall evaluation of CSMP at CEMREL.

5. Data Analysis

Average raw class scores were derived for each class for the vocabulary test, the computation category, and each of the six MANS process categories plus a total MANS score. For some tests, in order to increase the overall number of items without going beyond a reasonable time limit, a simplified form of item sampling was adhered to in the test booklet. A test was divided into halves with each half answered by a random half the class; in this case the average for each of the halves was added together to obtain a class score.

These class scores were then used as the unit for analysis in an Analysis of Covariance procedure comparing CSMP and non-CSMP classes, using class Vocabulary score as the covariate. At each grade level adjusted scores for the CSMP group and for the non-CSMP group were calculated, as well as the p-value on the analysis of covariance F-test, with degrees of freedom equal to total number classes minus three. Effect size was calculated by dividing the difference in adjusted means by the standard deviation of the control (non-CSMP) group.

C. Description of Results

Claim 1: "CSMP students are better able than comparable non-CSMP students to apply the mathematics they have learned to new problem situations".

Table 3 summarizes effect sizes by grade. All differences favored CSMP classes and were statistically significant at the .01 level except Producing Multiple Answers at fourth grade–the only result in which non-CSMP classes had a (marginally) higher score. Appendix B provides detailed data for each grade level.

Table 3
Summary of Effect Sizes, by Grade Level

	2nd	3rd	4th	5th	6th
Estimation	1.11	1.21	.94	.69	.52
Mental Arithmetic	1.96	1.07	1.13	.72	.52
Representations of Numbers	1.30	.77	.50	.44	.48
Number Patterns/Relationships	2.13	1.21	2.03	.85	.80
Word Problems	.89	.76	.45	.75	.44
Producing Multiple Answers	.69	.55	.02[1]	.37	.83
Total MANS	1.81	1.09	1.07	.71	.67

[1]In favor of non-CSMP classes

Claim 2. "CSMP students perform in traditional computation skills as well as comparable non-CSMP students"

Table 4 summarizes Computation scores for each grade level. CSMP classes had higher scores in three of the five grade levels. Only at second grade did the difference in adjusted means exceed one point or did the effect size exceed one-third of a standard deviation, and it was in favor of CSMP.

Table 4
Comparison of Class Computation Scores, by Grade Level

Grade	Raw Score Means		Standard Dev.		Adjusted Means		Effect	In Favor of	P-Value
	CSMP	non-CSMP	CSMP	non-CSMP	CSMP	non-CSMP	Size		
2	14.0	12.2	1.8	1.5	14.0	12.1	+1.30	CSMP	.01
3	17.5	16.4	2.1	1.8	17.3	16.9	+ .21	CSMP	.37
4	21.2	21.2	2.3	1.9	21.0	21.3	- .16	non-CSMP	.54
5	23.3	23.6	2.6	2.8	23.0	23.9	- .32	non-CSMP	.04
6	26.4	25.2	2.1	2.3	26.1	25.4	- .31	CSMP	.06

Evidence is presented below regarding three potential threats to the validity of these findings.

1. Potential effect for only high ability levels. The majority of classes tested had average vocabulary scores corresponding to the second quartile (50th to 75th percentile). A separate analysis of covariance was performed for only those classes with percentile rank below 50. Adjusted Total MANS scores are shown in Table 5. The differences in favor of CSMP for these classes were very consistent with differences derived from the entire set of comparison classes and the effect sizes were systematically higher. Thus, the overall higher scores for CSMP classes are not exclusively due to effects on higher ability classes.
[Claim 2, Computation. For this subgroup, CSMP classes had higher Computation scores than non-CSMP classes in four of the five grade levels, including the only difference that was statistically significant. See Appendix B.]

Table 5
Total MANS: CSMP versus non-CSMP for Classes
with Vocabulary Score Below 50th Percentile

Grade	Number of Classes CSMP	Number of Classes non-CSMP	Total MANS Average Raw Score CSMP	Total MANS Average Raw Score non-CSMP	Total MANS Adjusted Score CSMP	Total MANS Adjusted Score non-CSMP	CSMP Margin These Classes	CSMP Margin: All Comparison Classes
2	8	3	83.1	64.4	83.4	63.6	19.8	15.7
3	13	11	92.3	77.7	92.6	77.3	15.3	15.8
4	6	4	106.7	93.0	107.9	91.2	16.7	17.2
5	9	12	103.8	91.4	102.0	92.7	9.3	12.9
6	3	8	130.4	105.2	128.2	106.0	22.2	16.5

2. Potential Hawthorne effect. Most of the CSMP teachers participating in the comparison study were using the CSMP materials for the first time, creating a potential Hawthorne or enthusiasm effect. In six of the ten districts, some of these same teachers had their classes tested in a subsequent year. Table 6 shows that in these subsequent years adjusted Total MANS scores (derived from a separate analysis of covariance) were at least as high as in the comparison year. Thus, the evidence does not support a Hawthorne effect.
[Claim 2, Computation. At grades 2-3 comparison year classes had higher computation scores; at grades 4-6 later year classes had higher scores. None of the differences was statistically significant. See Appendix B.]

Table 6
Total MANS: Comparison Year versus Later Year for Same CSMP Teacher

Grade	Number of Teachers	Total MANS Average Raw Score Comparison Yr	Total MANS Average Raw Score Later Yr[1]	Total MANS Adjusted Score Comparison Yr	Total MANS Adjusted Score Later Yr[1]	In Favor of:
2	13	93.7	96.2	94.4	95.5	Later Year
3	16	102.6	107.4	104.1	105.9	Later Year
4	4	135.5	148.1	135.8	147.8	Later Year
5	12	138.7	138.7	138.7	138.7	
6	8	165.0	176.3	165.4	175.8	Later Year

[1]For grades 2 - 3, "later" = 1st year after the comparison study
For grades 4 - 6, "later" = 2nd or 3rd year after comparison study

3. Potential selectivity of districts participating in comparison studies. Comparison data reported in Tables 3 and 4 were derived from only those districts which did a formal CSMP - non-CSMP comparison study at that grade level. During the same time period, some other districts tested only their CSMP classes. Table 7 shows that these "other" CSMP classes had slightly higher adjusted Total MANS scores (using a separate analysis of covariance) than CSMP classes participating in the comparison studies. This supports the view that the performance of the comparison classes was not atypical of CSMP classes in general.
[Claim 2, Computation. At four of the five grade levels, these "other" CSMP classes had higher computation scores including the only differences which were statistically significant. See Appendix B].

Table 7
Total MANS: CSMP Classes Participating in Comparison
Studies Versus Other CSMP Classes Tested, 1984-91

Grade	Number of Classes		Total MANS Average Raw Score		Total MANS Adjusted Score		In Favor of:
	Comparison	Other	Comp. Classes	Other	Comp. Classes	Other	
2	41	50	33.9	35.3	95.7	98.6	Other
3	50	25	35.4	37.6	108.5	110.7	Other
4	36	82	29.1	31.0	139.3	146.1	Other
5	46	9	32.4	36.3	135.3	136.2	Other
6	43	10	27.3	30.5	168.5	169.8	Other

D. Summary of Supplemental Evidence

Comparisons with previous results. A previous submission of the CSMP program, approved by the Joint Dissemination Review Panel, 1984, reported similar data for the same claims. In that submission, class-level effect sizes were presented for the sixth grade. The present results are remarkably consistent with those of the previous submission, as is indicated below in Table 8.

Table 8
Comparison of Effect Sizes, Sixth Grade:
Present Data (1984-91) versus Previous JDRP Data (1984)

	Present Effect Size	Previous Effect Size
Estimation	.52	.41
Mental Arithmetic	.52	.63
Number Representations	.48	.38
Relations/Number Patterns	.80	1.00
Word Problems	.44	.56
Multiple Answers	.83	.91
Total MANS	.67	.63

Student level effect size from the previous submission was .37, but could not be calculated for 1984-91 data because some student level data was not retained. However, because of the similarity in findings illustrated in Table 8, a similar figure can be inferred. In any case, student level differences are large enough that separate norms tables are used when reporting student scores. For example, for sixth grade students, a Total MANS raw score corresponding to the 50th CSMP percentile rank would correspond to the 61st non-CSMP percentile rank. Since sixth grade results showed the smallest gains by CSMP it can be assumed that effect sizes and percentile rank differences are higher at other grades.

Some participating districts have collected data from one or more standardized tests. These data collection efforts are incompatible from site to site, but districts have consistently reported higher achievement by their CSMP students on the Applications or Problem Solving sections of these tests, mirroring the findings in this submission on the MANS Tests. Districts have reported very inconsistent findings with regard to the Computation section of these tests, with about equal numbers reporting no difference, differences in favor of CSMP, and differences in favor of non-CSMP, again mirroring the data in this submission that there is no consistent pattern of CSMP students' performance in computation being either better or worse than that of non-CSMP students.

E. Interpretation and Discussion of Results

1. Relationship Between Effect and Treatment

Claim 1. Results from the MANS process categories indicate overwhelmingly that CSMP classes at every grade level and ability level perform better in those processes than non-CSMP classes. Furthermore, the data are consistent with curriculum comparison data obtained in earlier years from different school districts. The CSMP curriculum contains many problem situations and extensive teacher lesson plans which give students experience in the kinds of thinking processes covered by the MANS Tests - though not using the same language or problems as are in the tests. The spiral nature of the curriculum allows these processes to occur and recur frequently in the curriculum and in many different guises. Hence, this claim is consistent with the nature of the materials in the CSMP curriculum.

Claim 2. Results on Computation tests were mixed, favoring CSMP or non-CSMP at different grade levels. The only significant difference was in favor of CSMP in second grade. The CSMP curriculum does not particularly emphasize rote computation and skill in algorithms per se, but does provide many opportunities for practice of mental arithmetic and estimation skills embedded in other mathematical activities. The evidence indicates that the curriculum develops computation skills comparable to other mathematics curricula.

2. Control of Rival Hypotheses

Since the design was based on a comparison of intact classes, all tested at the end of the year, factors such as testing, maturation, attrition, and differential selection of groups are controlled in the design. Teacher effects were controlled in part by having the same teacher's classes tested - one year with the regular district curriculum and the next year with CSMP. Generally, teacher effects would have most force at the earliest grades when the original selection of teachers occurred. In later years, as CSMP was implemented into successively higher grades, the CSMP teacher more or less inherited the class from previous teachers. CSMP and non-CSMP classes were of roughly equivalent ability as measured by Vocabulary scores and differences were controlled statistically in the analysis by using Vocabulary score as a covariate. Potential Hawthorne effects, selectivity of sites tested, and different results by class ability levels were all investigated; the data do not support these rival hypotheses.

F. Educational Significance of Results

Educational significance is always difficult to assess. Two factors are important in this submission. First, the MANS Tests are focussed on applications of mathematics to situations which are relatively unfamiliar to the students being tested and require higher level cognitive activity than merely demonstrating a skill or learned content. Second, improvement on these kinds of measures has been notoriously difficult to achieve.

For example, the 40-point decline in the Mathematics section of the SAT observed from 1963 to 1970 is equivalent to slightly less than 1/2 raw score standard deviation. Also, the "most salient finding" reported by the 1983 National Assessment of Educational Progress in mathematics was that "13 year olds have improved dramatically" (the improvement was about 3 percentage points) and that "of particular significance is the 8 percentage point gain for 13 year olds in heavily minority schools." These improvements, considered important by educators, are comparable to or smaller than those reported in this submission.

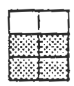
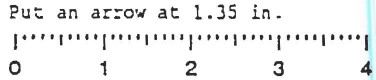
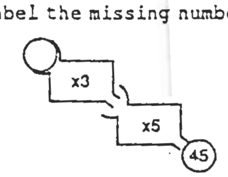
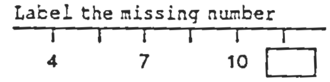
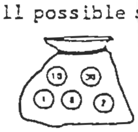
The same NAEP reports allude to the difficulty in improving students' abilities to apply mathematics: "Even the 13-year-olds, who made significant gains on routine problem solving, showed no change in their performance on non-routine problems."

Finally, in the NAEP discussion of the major implications of the findings:

> "Schools are doing a good job of teaching mathematical topics that are relatively easy to teach... there was very little change in topics that are relatively difficult to teach, such as non-routine problem solving... Changes at the higher cognitive levels will occur only when higher level cognitive activity becomes a curricular and instructional focus."

APPENDIX A

ABBREVIATED SAMPLE MANS ITEMS

Category	Second Grade	Fourth Grade	Sixth Grade
Estimation	90 - 12 is in which interval? 0 - 10 - 50 - 100 - 500	602 is about ? as large as 298? 2, or 5 or 10 times	8 1/2 - 8 is: <1 or =1 or >1 Which interval contains 1,002.5 - 21.5? 0 - 1 - 10 - 20 - 50 - 100
Mental Arithmetic (scratch word not allowed)	300 - ? = 250	Hit -> gain 5; Miss -> lose 1 Start with 10 below zero Number of hits = 1 Number of misses = ? End with 12 below zero	12 X 75 = 900 13 X 75 = ? 1/2 X ? = 40
Number Represen-tations	Write "two thousand, eleven" 100 more than 901 is ?	How much is shaded?  1/2 1/4 2/3 3/4 Name the 2nd largest 3-digit number using only 2, 5, 7, 8	Which are equivalent to 1/3? 2/6 11/31 3/15 4/12 50/150 Put an arrow at 1.35 in. 
Number Patterns/ Relations	20 - 10 8 - 4 10 - 5 6 - ? Which is larger? 585 + 250 or 580 + 290	What are the missing numbers? ?, 50, ?, 200, 400, 800 Label the missing number 	Which is larger? 5/2 or 5/4 0.9 or 0.11111 Name a fraction (decimal) that is: larger than 1/3 but smaller than 1/8 larger than 0.2 but smaller than 0.3 Label the missing number 
Word Problems	One stage problems, requiring easy computations, with pictorial format and read by the tester	Two-stage Miscellaneous data Word problem approximations	Three stage Word problems with fractions, decimals Novel word problems
Producing Multiple Answers	Write number sentences about 8 8 = 9 - 1	Take out 2 balls together Add to get a total score Give all possible scores 	Use only even numbers They must be divisible by 10 They must be smaller than 80 Give all possible numbers

APPENDIX B

SUPPORTING DATA

Note. In Tables A1 - A5, all differences are in favor of CSMP unless otherwise noted.

Table A1
Detailed Comparison Data, Second Grade

MANS Category	Raw Score Means		Standard Dev.		Adjusted Means		P-	Effect
	CSMP	non-CSMP	CSMP	non-CSMP	CSMP	non-CSMP	Value	Size
Estimation	10.0	8.5	2.1	1.7	10.1	8.3	.01	1.11
Mental Arithmetic	20.3	15.1	4.1	3.0	20.6	14.7	.01	1.96
Representations of Numbers	17.5	15.8	2.2	1.5	17.6	15.7	.01	1.30
Numb Patterns/Relationships	29.1	23.8	5.1	2.8	29.3	23.4	.01	2.13
Word Problems	6.2	5.6	1.0	1.0	6.3	5.5	.01	.89
Producing Multiple Answers	10.6	9.4	2.0	2.0	10.6	9.3	.01	.69
Total MANS	93.7	78.1	15.1	9.8	94.5	76.8	.01	1.81

Table A2
Detailed Comparison Data, Third Grade

MANS Category	Raw Score Means		Standard Dev.		Adjusted Means		P-	Effect
	CSMP	non-CSMP	CSMP	non-CSMP	CSMP	non-CSMP	Value	Size
Estimation	22.0	17.9	3.0	2.5	21.7	18.6	.01	1.21
Mental Arithmetic	17.7	13.4	3.0	3.0	17.4	14.2	.01	1.07
Representations of Numbers	9.8	8.1	1.5	1.5	9.6	8.5	.01	.77
Numb Patterns/Relationships	31.7	24.2	4.6	4.7	31.1	25.4	.01	1.21
Word Problems	8.3	7.0	1.4	1.1	8.2	7.3	.01	.76
Producing Multiple Answers	17.2	14.8	2.3	2.8	16.9	15.4	.01	.55
Total MANS	106.8	85.5	14.8	14.3	104.9	89.4	.01	1.09

Table A3
Detailed Comparison Data, Fourth Grade

MANS Category	Raw Score Means		Standard Dev.		Adjusted Means		P-	Effect
	CSMP	non-CSMP	CSMP	non-CSMP	CSMP	non-CSMP	Value	Size
Estimation	30.0	26.5	3.3	2.8	29.6	26.9	.01	.94
Mental Arithmetic	22.0	17.4	3.6	3.3	21.6	17.9	.01	1.13
Representations of Numbers	17.9	16.3	2.6	2.0	17.6	16.6	.01	.50
Numb Patterns/Relationships	34.6	24.7	5.1	4.2	34.0	25.4	.01	2.03
Word Problems	13.7	12.1	2.5	2.1	13.4	12.5	.01	.45
Producing Multiple Answers	18.1	17.5	2.5	2.7	17.8	17.8	.88[1]	.02[1]
Total MANS	136.3	114.5	18.5	15.6	133.9	117.1	.01	1.07

[1]In favor of non-CSMP classes

Table A4
Detailed Comparison Data, Fifth Grade

MANS Category	Raw Score Means		Standard Dev.		Adjusted Means		P-	Effect
	CSMP	non-CSMP	CSMP	non-CSMP	CSMP	non-CSMP	Value	Size
Estimation	16.3	14.8	1.7	1.7	16.1	15.0	.01	.69
Mental Arithmetic	23.5	20.0	3.5	3.6	23.1	20.4	.01	.72
Representations of Numbers	23.2	21.0	3.4	3.5	22.9	21.4	.01	.44
Numb Patterns/Relationships	33.3	27.7	4.9	5.1	32.7	28.3	.01	.85
Word Problems	12.2	10.2	2.5	2.0	11.9	10.5	.01	.75
Producing Multiple Answers	23.3	20.7	4.2	4.7	22.9	21.2	.01	.37
Total MANS	131.8	114.4	19.1	18.1	129.6	116.7	.01	.71

Table A5
Detailed Comparison Data, Sixth Grade

MANS Category	Raw Score Means		Standard Dev.		Adjusted Means		P-	Effect
	CSMP	non-CSMP	CSMP	non-CSMP	CSMP	non-CSMP	Value	Size
Estimation	22.9	20.3	3.2	3.5	22.5	20.7	.01	.52
Mental Arithmetic	21.6	18.5	3.5	4.2	21.1	19.0	.01	.52
Representations of Numbers	27.6	24.9	3.7	3.7	27.1	25.4	.01	.48
Numb Patterns/Relationships	46.2	39.3	5.5	6.5	45.3	40.1	.01	.80
Word Problems	13.5	11.7	2.2	2.6	13.1	12.0	.01	.44
Producing Multiple Answers	33.3	27.7	4.2	5.2	32.6	28.3	.01	.83
Total MANS	165.1	142.3	21.1	24.3	161.8	145.4	.01	.67

Table A6
Adjusted Mean Computation Scores
Corresponding to Tables 5-7

Grade	Classes Below 50th Vocabulary Percentile (See Table 5)		Comparison with Later Years Same CSMP Teacher (See Table 6)		CSMP Comparison Classes versus CSMP Classes in Other Districts (See Table 7)	
	CSMP	non-CSMP	Comparison Year	Later Year	Comparison Classes	Other CSMP Classes
2	13.2	10.1[1]	14.7	14.0	14.2	14.2
3	16.2	15.0	17.6	17.2	17.6	18.6[1]
4	20.3	19.2	21.6	22.1	21.6	22.8[1]
5	20.0	20.9	23.2	23.8	23.6	25.2[1]
6	24.6	22.3	26.1	26.3	26.6	26.4

[1]Significant at .05 level of significance