Chapter 3: Method

This chapter describes the research methodology used to investigate the joint effects of a semestered block schedule and the IMP curriculum on student mathematics learning at Suburban High School. First, the chapter provides a general description of the school community, including the students and teachers at the school. Then, the chapter describes the traditional curriculum and schedule that had previously been used at Suburban High School, and the IMP curriculum/semestered block schedule that were adopted. Then, the chapter describes data collection and statistical methodology for each of three approaches used to compare mathematics learning of students using the traditional schedule and curriculum to that of students using the block schedule and IMP curriculum:

1. analysis of an Algebra Achievement Test completed by students at the end of Grade 11,
2. analysis of student enrollment in mathematics courses as reported by student transcripts, and
3. analysis of student participation in and exam grades for Advanced Placement (AP) mathematics classes as reported on yearly School Profiles published by Suburban High School.

The chapter ends with a short description of qualitative data collected to help describe unique aspects of how the IMP curriculum and semestered block schedule were implemented at Suburban High School.

Participating Students

Participants in this study's analysis of algebra achievement include students at Suburban High School who were enrolled in the eleventh grade during the spring of 1997 (Traditional cohort), the spring of 2000 (First Reform cohort), or the spring of 2001 (Second Reform cohort). Participants in this study's transcript analysis include those who participated in the analysis of algebra achievement, plus students who used a traditional schedule and curriculum throughout high school, and graduated in the spring of 1996 or the spring of 1997. The analysis of Advanced Placement (AP) participation used information from additional years, analyzing transcripts of students who participated in other parts of the study, plus those of all other students who graduated between the spring of 1995 and the spring of 2001.

Suburban High School is the only high school in a small, relatively affluent school district in the Eastern United States. A number of colleges are located nearby, and the local population tends to be well educated. Traditionally, more than 75% of students at Suburban High School attend 4-year colleges after graduating from high school; counting both 2- and 4-year colleges, about 90% of Suburban High School students go on to college immediately after high school graduation.

The school generally enrolls between 900 and 1000 students each year, distributed as 200+ per grade, in Grades 9 through 12. Most students at Suburban High School have graduated from the district's only public middle school, which shares a campus with the high school.

In 1997, the Department of Education in the state where Suburban High School is located rated the school as being in the top quintile (20%) statewide in socio-economic status. In 2000 and 2001, the school was rated in the second-from-top quintile.

Since the 1991-92 school year, when students in the Traditional cohort were in sixth grade, students at the feeder middle school have used the *Visual Mathematics* curriculum (Foreman & Bennett, 1991). *Visual Mathematics* is an innovative curriculum designed to reflect the reforms characterized in the NCTM *Standards*. Published by a small non-profit company, *Visual Mathematics* is currently used only in a handful of districts across the United States. Because of their unusual middle school experience, students in both the Traditional cohort and students in the First Reform and Second Reform cohorts entered high school with a unique prior mathematics experience that may influence their readiness for solving the kind of problems emphasized by *Standards*-based curricula.

Participating Teachers

Mobility among mathematics teachers at Suburban High School is limited, so teachers are relatively experienced. Also, Mrs. Sullivan, who served jointly from 1984 through the spring of 1998 as mathematics supervisor for the school district and as high school mathematics department chair, had advocated adoption of both *Visual Mathematics* (Foreman & Bennett, 1991) at the middle school and the *Integrated Mathematics Program* (Fendel, Resek, Alper, & Fraser, 1997) at the high school. During that time, Mrs. Sullivan recruited teachers whose philosophy was broadly compatible with the teaching methods used in these two curricula.

Professional Development

Even before their school adopted the IMP, teachers at Suburban High School had received extensive professional development in problem-centered instruction. Each summer from 1993 through 1995, nearly all mathematics teachers from both the feeder middle school and the high school participated in 30 hours of professional development centering around the Visual Mathematics middle school curriculum. Once they had adopted the IMP curriculum, teachers at the high school received additional professional development to that curriculum. In general, teachers attended 30 hours of professional development on utilizing the IMP curriculum each summer for four consecutive

summers, beginning with a course in 1996 focusing on the first IMP textbook, and ending with a course in 1999 focusing on the fourth IMP textbook. Some teachers attended additional short workshops during the school year. Teachers who joined the mathematics faculty subsequent to 1996 also have enrolled in 30-hour courses focusing on each of the IMP textbooks, beginning with Book 1.

Planning Time

Under the traditional schedule, teachers at Suburban High School had available one 43-minute planning period per day. Under the semestered block schedule this was increased to a planning period consisting of one 80-minute block per day.

Traditional Schedule

Students in the Traditional cohort were enrolled in a 7-period day using a traditional mathematics curriculum throughout Grades 9-11. Each year students were enrolled in seven courses at a time, with each course meeting 43 minutes per day for the entire 180-day school year.

During the 1997-98 school year, when students in the Traditional cohort were in twelfth grade, Suburban High School adopted a block schedule school-wide. This study refers to the schedule adopted in 1997-98 as the "Pilot" block schedule, to distinguish it from a slightly different form of block schedule that was used in subsequent years. In 1997-98 students took 4 courses at a time, each course meeting 80 minutes per day over an 80-day semester. A 20-day spring session was reserved for special in-depth projects. This schedule change did not affect students in the Traditional cohort prior to the end of Grade 11, when they completed the Algebra Achievement test. However, the new schedule did affect their course-taking in twelfth grade. For this reason, when comparing

student course-taking under a traditional schedule and curriculum to course-taking under a block schedule and the IMP curriculum, this study used transcript data from the previous two cohorts of Suburban High School students: those who graduated in the spring of 1997, and those who graduated in the spring of 1996.

Traditional Curriculum

Suburban High School offered students in the Traditional cohort three "levels" of courses: Honors, College Preparatory (CP), and Academic Assisted (AA). For students taking Honors or College Preparatory courses, the core mathematics sequence was Algebra 1, Geometry, and Algebra 2. For Algebra I, they used *Algebra 1* (Larson, Kanold, & Stiff, 1995), published by D.C. Heath. For Geometry, they used *Geometry for Enjoyment and Challenge* (Rhoad, Malauskas, & Whipple, 1991), published by McDougal/Littel. For Algebra II, they used *Algebra 2* (Larson, Kanold, & Stiff, 1993), published by D.C. Heath. After completing the three core courses, students could enroll in Contemporary Mathematics, Algebra 3/Trigonometry, Functional Analysis, Discrete Analysis, Statistics, Calculus A/B and Calculus B/C. There were two distinctions between Honors and College Preparatory courses: 1) Honors students generally were given more challenging problem sets within any given unit; 2) Honors courses tended to be faster paced, covering a few extra concepts. For example, the Honors Algebra 2 class covered basic trigonometric functions, as well as arithmetic and geometric sequences, whereas the College Preparatory version of the same course did not.

For students in Academic Assisted courses, the first two core mathematics courses used texts entitled *Math Matters: An Integrated Approach* (Lynch & Olmstead, 1993a, 1993b) published by Southwestern. For the third course in the core sequence,

students could either enroll in a course using the third book of *Math Matters: An Integrated Approach* (Ebos & Zolis, 1987) or else enroll in a low level Algebra 2 course called "Algebra 2 Career/College Prep", using the texts *Algebra* (McConnell, 1993) and *Geometry* (Coxford, 1991), published by Scott Foresman. Academic Assisted students who continued their math studies in twelfth grade could enroll in Contemporary Mathematics or Algebra 3/Trigonometry.

Reform Cohorts

Students in the Reform cohorts used a semestered block schedule and the IMP curriculum. The First and Second Reform cohorts were, respectively, the second and third group of students at Suburban High School to use the new schedule and curriculum.

The Semestered Block Schedule

The school piloted a semestered block schedule with ninth graders during the 1996-97 school year—the year before the students in the First Reform cohort entered high school. As noted above, during the 1997-98 school year when students in the First Reform cohort were in ninth grade, Suburban High School implemented a block scheduling school-wide. That year, students took 4 courses at a time, each course meeting 80 minutes per day for one 80-day semester. A 20-day spring session was reserved for special in-depth projects.

During the 1998-99 school year, when students in the First Reform cohort were in tenth grade and students in the Second Reform cohort were in ninth grade, the schedule was modified to consist of two 90-day semesters, as the 20-day spring session was eliminated. The schedule has remained unchanged since that time.

The IMP Curriculum

At the same time Suburban High School began piloting a semestered block schedule, the school also began piloting the IMP curriculum. Each of IMP's four year-

long textbooks contains five modules. Individual modules are usually designed around a single over-arching problem whose solution requires a number of key mathematics concepts. Students spend several weeks working on sub-problems and related problems, developing the mathematics skills and knowledge needed to solve the module's central problem.

The four IMP textbooks cover most of the material contained in the traditional 3-year sequence of algebra 1, geometry, and algebra 2, plus some additional material generally contained in a trigonometry/pre-calculus course. In addition, units dealing with matrix algebra and/or units dealing with probability and statistics are included in each of the four textbooks.

All regular education students in the Reform cohorts used a sequence of four IMP courses to replace the traditional three-course core sequence. The IMP courses replaced Algebra 1, Geometry, and Algebra 2, at either the Honors, College Preparatory, or Academic Assisted level.

Suburban High School continues to distinguish among "levels" of courses. Different levels used the same IMP curriculum, but moved through it at differing speeds. The Honors level courses complete an entire IMP textbook (five modules) in each course; the College-Preparatory level courses finish four IMP modules per course, and the Academic-Assisted level courses finish three or four IMP modules per course. Table 1 displays the sequence each ability level followed as it completed the IMP modules as of 2000-2001. As Table 1 shows, Suburban High School adapted the IMP curriculum to the fewer hours available for instruction per course under a block schedule by completing less than one year's worth of material per course for all except Honors-level students.

Appendix E supplements Table 1 by providing a detailed description of the key concepts and skills within each IMP module as described on the Suburban High School 2000-2001 course syllabi.

It should be noted that, while the information in Table 1 is representative of the coursework students completed as they went through the Integrated Math sequence, the syllabi did change somewhat from year to year. In particular, in the earlier years of implementation teachers were less familiar with the IMP content, and course syllabi contained completed fewer modules per course than is reflected in the 2000-2001 data displayed in Table 1.

Also, teachers at Suburban High School have not always been able to complete the entire course syllabus. Partly because their students have the opportunity to study statistics either in an Advanced Placement or standard format after completing the IMP sequence, when pressed for time teachers have usually dropped one or more of the probability and statistics modules from their syllabus. According to teachers at the school, *The Game of Pig* and *The Pit and the Pendulum* have nearly always been taught as described in course syllabi, but *Is There Really a Difference?* and *Pennant Fever* have only sometimes been taught. *The Pollster's Dilemma* has generally not been taught, but it is the intent of the teachers that it will be in future years.

Table 1. Sequence of IMP modules completed by each ability group at Suburban High School

	Course in Which Module Was Completed		
	Honors	College Prep	Academic Assisted
IMP Textbook Year 1			
Patterns	Integrated Math 1	Integrated Math 1	Integrated Math 1

The Game of Pig	Integrated Math 1	Integrated Math 1	Integrated Math 1
The Overland Trail	Integrated Math 1	Integrated Math 1	Integrated Math 1
The Pit and the Pendulum	Integrated Math 1	Integrated Math 2	Integrated Math 2
Shadows	Integrated Math 1	Integrated Math 1	Integrated Math 2
IMP Textbook Year 2			
Solve It!	Integrated Math 2	Integrated Math 2	Integrated Math 2 and 3
Is There Really a Difference?	Integrated Math 2	Integrated Math 3	-
Do Bees Build it Best?	Integrated Math 2	Integrated Math 2	Integrated Math 3
Cookies	Integrated Math 2	Integrated Math 3	Integrated Math 3
All About Alice	Integrated Math 2	Integrated Math 2	Integrated Math 4
IMP Textbook Year 3			
Fireworks	Integrated Math 3	Integrated Math 3	Integrated Math 4
Orchard Hideout	Integrated Math 3	Integrated Math 3	Integrated Math 4
Meadows or Malls?	Integrated Math 3	-	Integrated Math 4
Small World, Isn't It?	Integrated Math 3	Integrated Math 4	-
Pennant Fever	Integrated Math 3	-	-
IMP Textbook Year 4			
High Dive	Integrated Math 4	Integrated Math 4	-
As the Cube Turns	Integrated Math 4	-	-
Know How	Integrated Math 4	Integrated Math 4	-
The World of Functions	Integrated Math 4	Integrated Math 4	-
The Pollster's Dilemma	Integrated Math 4	-	-

Because students in different ability groups completed differing numbers of modules per course, there were some IMP modules that students in lower ability groups do not cover until twelfth grade, and others they did not cover at all. In general, by the end of Integrated Math 3, College Preparatory students completed most of the material usually contained in Algebra 1, Geometry, and Algebra 2. By the end of Integrated Math 4, College Preparatory students completed much of the material generally contained in a Trigonometry/Pre-Calculus course as well. The four modules deleted from the College Preparatory curriculum dealt primarily with probability, statistics, and matrix algebra.

Two of the modules that weren't addressed until Integrated Math 4 in Academic Assisted classes deal with concepts contained in the algebra achievement test utilized by this study. *All About Alice* deals extensively with exponential functions, and *Fireworks* presents extensive opportunities to work with quadratic equations. A third module, *Orchard Hideout*, covers key geometry concepts, and the fourth, *Meadows or Malls?* covers matrix algebra concepts that were left out of the College Preparatory classes. The modules that Academic Assisted classes never cover include three of the four dealing with probability and statistics, one dealing with matrix algebra, and nearly all of the Trigonometry/Pre-Calculus content. Some Academic Assisted students who wished to study the Trigonometry/Pre-Calculus content did so by enrolling in Integrated Math 4 College Preparatory after completing Integrated Math 4 Academic Assisted.

It should be noted while the Algebra topics tested on the Algebra Achievement test used by this study are addressed by IMP before the end of the Year 3 textbook, there is some review and extension of Algebra concepts in IMP Year 4. This is particularly

true of quadratic equations, which are among the topics addressed in the modules *Know How* and *High Dive*. Honors students completed both modules as part of Integrated Math 4, while College Preparatory students completed *Know How* in Integrated Math 4, but did not complete *High Dive*. For this reason, students in the Reform cohorts who took Integrated Math 4 in their senior year had not completed all of their Algebra study at the time the Algebra Achievement test was administered.

After completing the four Integrated Math courses, students could take Contemporary Mathematics, Functional Analysis, Discrete Analysis, Statistics, Calculus A/B and Calculus B/C. The Algebra 3/Trigonometry course, which had contained a mixture of Algebra review and more advanced topics that were now studied in Integrated Mathematics 4 College Preparatory, was discontinued.

Schedule and Curriculum: Summary

Table 2 provides the timetable followed by Suburban High School for implementation of the IMP curriculum and the semestered block schedule.

Table 2. Implementation of Semestered Block Schedule and IMP Curriculum at Suburban High School

	1995-1996 and earlier		1996-1997		1997-1998	
Student Group	Curriculum	Schedule	Curriculum	Schedule	Curriculum	Schedule
Grade 9 CP/AA	Traditional	Traditional	IMP	Block[a]	IMP	Block[a]
Grade 9 Honors	Traditional	Traditional	Traditional	Block[a]	IMP	Block[a]
Grade 10 CP/AA	Traditional	Traditional	Traditional	Traditional	IMP	Block[a]
Grade 10 Honors	Traditional	Traditional	Traditional	Traditional	Traditional	Block[a]
Grade 11 CP/AA	Traditional	Traditional	Traditional	Traditional	Traditional	Block[a]
Grade 11 Honors	Traditional	Traditional	Traditional	Traditional	Traditional	Block[a]
Grade 12 CP/AA	Traditional	Traditional	Traditional	Traditional	Traditional	Block[a]
Grade 12 Honors	Traditional	Traditional	Traditional	Traditional	Traditional	Block[a]

	1998-1999		1999-2000		2000-2001 and later	
Student Group	Curriculum	Schedule	Curriculum	Schedule	Curriculum	Schedule
Grade 9 CP/AA	IMP	Block	IMP	Block	IMP	Block
Grade 9 honors	IMP	Block	IMP	Block	IMP	Block
Grade 10 CP/AA	IMP	Block	IMP	Block	IMP	Block
Grade 10 honors	IMP	Block	IMP	Block	IMP	Block
Grade 11 CP/AA	IMP	Block	IMP	Block	IMP	Block
Grade 11 honors	Traditional	Block	IMP	Block	IMP	Block
Grade 12 CP/AA	Traditional	Block	IMP	Block	IMP	Block
Grade 12 honors	Traditional	Block	Traditional	Block	IMP	Block

[a] In 1996-7 and 1997-8, mathematics instruction under the block schedule was conducted during two 80-day semesters. A 20-day spring semester was reserved for special-interest courses. In 1998-9 and thereafter, mathematics instruction under the block schedule was conducted during two 90-day semesters. The 20-day spring semester was discontinued.

Data Collection

This study used five primary sources of data from Suburban High School. First, it analyzed results of an Algebra Achievement test completed by eleventh graders in one Traditional cohort and two Reform cohorts. Second, it used student scores from a sixth grade test administered by the Educational Records Bureau as a covariate. Third, it analyzed transcripts from an automated data base containing information from the spring of 1991 through the spring of 2001. Fourth, it analyzed documents provided by the school, including course syllabi and annual school profiles. Fifth, it analyzed information provided in conversations with key informants at the school.

Algebra Achievement Test

This study used a 3-part Algebra Achievement test designed by the Core-Plus Mathematics Project. Part 1 emphasized the type of contextualized problem solving that is typical of Core-Plus, IMP, and other reform curricula. Part 2 emphasized problems typical of traditional mathematics curricula: context-free symbolic manipulations that call for transformation of algebraic expressions and solutions of equations and systems. Items in Part 2 were adapted from released ACT examinations and from items that commonly appeared on college placement tests. Part 3 required collaborative work on a single extensive open-ended problem and was designed to be completed by students in pairs. The Algebra Achievement test was intended to be administered at the end of Grade 11 and focuses on algebra topics that are generally completed by that time.

The Algebra Achievement test designed by Core-Plus has several advantages. Like the IMP, Core-Plus is a curriculum developed under a National Science Foundation grant to implement the NCTM *Standards* at the high school level. The Algebra Achievement test was designed specifically to fulfill the purpose of the proposed study: to compare the effects of a *Standards* -based curriculum to those of more conventional curricula. Since this study compares learning under the IMP curriculum to learning under a more traditional curriculum, it is important to use a test that is fair to both. The Algebra

Achievement test accomplishes this, by measuring both the kind of problem solving and applications emphasized by the NCTM *Standards*, as well as more traditional mathematics skills. Further, there is no chance that the test was unconsciously "tailored" to favor either the IMP or the Traditional curriculum, as this test was not designed by the researchers in this study or by anyone involved with either curriculum.

In order to sample a wide variety of problems, the Core-Plus researchers designed four parallel forms for Part 1, two parallel forms for Part 2, and three parallel forms for part 3. They administered the test via matrix sampling; that is, each student was randomly given one form for each of the three parts of the test. However, matrix sampling was not feasible at Suburban High School, given both the smaller sample size and the desire of Suburban High School teachers to maintain a simple testing program so results could be easily explained to the community. Therefore, this study used one form for each part of the test, selected by teachers at Suburban High: Part 1, form C; Part 2, form A; and Part 3, form A. As noted by the test authors (Huntley, et al., 2000), scores across forms of this test tend to be consistent, so the decision to use only one form was expected have little negative impact on the validity of results at Suburban High. The three parts of the Algebra Achievement test used in this study are contained in Appendix A.

Testing in the spring of 1997: Traditional cohort. In the spring of 1997, Suburban High School students in the Traditional cohort were in eleventh grade and nearly all of them were enrolled in mathematics. They completed the three parts of the Algebra Achievement test in mathematics class during two days in May 1997. On the first day of testing, individual students completed Part 1 of testing. On the second day of testing, individual students completed Part 2. Then, students within classrooms chose partners and together these pairs completed Part 3 of the test. Suburban High School mathematics teachers conducted the 1997 testing and archived the results so it would be possible in later years to compare the achievement of students who had studied under the

new curriculum and schedule to that of the 1997 eleventh graders, who had studied under a traditional curriculum and schedule. In 1997, 89.9% of eligible eleventh graders participated in at least one day of testing. Some of the students who missed the test were unable to participate because of school-scheduled extracurricular activities, and others did not participate due to absence.

Pilot Testing in the spring of 1999. During two days in May 1999, teachers at Suburban High School administered the Core-Plus Algebra test to eleventh graders school-wide. A pilot study compared results of this assessment to those of the May 1997 assessment. Lessons learned from the pilot study indicated that a number of steps needed to be taken to ensure that future testing conditions would be as close as possible to what they had been in 1997. Specifically, in 1999 many students were administered the test in settings that did not resemble a mathematics class, proctored by a non-mathematics teacher who did not create a serious atmosphere. Often, calculators were not available when they should have been. These problems were corrected in the spring of 2000, when the testing to be used for this proposed study was conducted.

Testing in the spring of 2000: First Reform cohort. Suburban High School students in the First Reform cohort completed the three parts of the Core-Plus Algebra test during two days in May 2000, when they were in eleventh grade. Because of the semestered block schedule, many eleventh graders were not enrolled in mathematics during this spring semester. Therefore, for the one hour needed each day for test administration, eleventh graders moved to a mathematics classroom or other classroom proctored by a mathematics teachers—or, in a few cases, by a science teacher. Since all students were enrolled in English during the second semester of eleventh grade, the classroom to which students reported was determined by their English class. In 2000, 90.4% of eligible eleventh graders participated in at least one day of testing. As before,

some of the students who missed the test were unable to participate because of school-scheduled extracurricular activities, and others did not participate due to absence.

Observers reported that the atmosphere and testing conditions in 2000 were very similar to what they had been in 1997. However, discussions after the testing raised concern about the way students were assigned to pairs during the second day of test administration. As in 1997, individual students completed Part 1 on the first day of testing and Part 2 at the beginning of the second day of testing. Then, students within classrooms chose partners and together these pairs completed Part 3 of the test. However, in 2000 students were tested within English class groupings, so it was likely that many pairs consisted of students who had completed differing levels of mathematics. This contrasted with the situation in 1997, when students were tested within a mathematics class, and so automatically paired with another student who had completed the same level of mathematics. Testing conditions in 2001 were adjusted to correct this potential problem.

Testing in the spring of 2001: Second Reform cohort. Suburban High School students in the Second Reform cohort completed the three parts of the Core-Plus Algebra test during two days in May 2001, when they were in eleventh grade. Testing conditions were the same as in the spring of 2000, with two exceptions.

First, students were given class credit for showing up at the test. This change was intended to increase the participation rate, and may have been marginally successful in doing so. In 2001 91.4% of eligible juniors participated in at least one day of testing.

Second, when administering Part 3, teachers requested that when choosing partners, students select someone whose most recent mathematics course was the same

level as their own. This was intended to make testing conditions more similar to what they had been in 1997.

Scoring procedures. The Algebra Achievement test contained open-ended questions that needed be scored using a rubric. For this study, a number of changes were made to the rubric used by the original designers of the test, so that the rubric would be easier to use validly and reliably. The most important change was the selection of anchor papers and practice papers, keyed to each possible score for each item in Parts 1, 2, and 3. In almost all cases, anchor papers and practice papers were selected from actual student responses to earlier administrations of the assessment that had been conducted by Core-Plus researchers. In the few instances where no student paper exemplified a particular response covered by the rubric, this researcher developed an appropriate "anchor paper". Procedures for training scorers for this study were based on professional standards used for the National Assessment of Educational Progress (NAEP), as described by Bourgeacq, et al. (1997). Appendix B contains the rubric used for scoring, and appendices C and D contain the anchor and practice papers used in training.

Because Part 3 is the most difficult section to score, each student submission of Part 3 was reviewed by three independent raters. The raters used a scale of 0 to 4. Two independent raters scored each question on Part 1 and Part 2 of the test. In cases of disagreement, raters reached consensus by discussion and persuasion, not voting. Part 1 and Part 3 were scored by an expert panel of college mathematics professors and retired high school mathematics teachers. Because Part 2 was relatively easier to grade, it was scored by two undergraduate mathematics majors.

Scoring was accomplished at two separate times. In the winter of 1999-2000, all tests from the Traditional cohort and the Pilot cohort were scored, as part of the pilot study. Then, in the summer and fall of 2001, Part 2 and Part 3 of the tests from 1997 were re-scored, and Part 1, Part 2, and Part 3 of the tests completed in 2000 or 2001 were scored. The 1997 Part 2 tests were re-scored because it proved impossible to get the same individuals who had scored Part 2 tests for the pilot study to complete the scoring in 2001, and it was deemed important to have the same raters for tests completed by students in the Traditional cohort and tests completed by students in the Reform cohorts.

Scorers who had graded Part 1 and Part 3 of the test in the pilot study were available to complete scoring in 2001. Before beginning to score the new tests, the scorers completed a "drift test" by re-scoring 20 tests that they had scored during the pilot study. The 20 tests to be re-scored were mixed in with 20 new tests, so that the scorers were more or less blind as to whether they were re-scoring an old test or scoring a new test.

The drift test found that there may have been a systematic difference between original scores and re-scores for Part 3. The consensus score was higher on the re-score than on the original score for 3 of the 20 tests, and the same on the re-score as on the original score for remaining 17 of the 20 tests. With the re-score being higher 15% of the time, it was possible that scorers had drifted towards scoring more leniently. The first attempted to correct this was by re-training. However, a second drift test found that the Part 3 scorers had overcompensated, consistently scoring more harshly on the re-score. Therefore, it was decided to mix the 1997, 2000, and 2001 Part 3 tests together, and have the raters score tests from all three cohorts at once. This new set of scores was used for

all analyses in this study; scores on Part 2 and Part 3 that had been computed during the pilot study were discarded.

In contrast to results for Part 3, results from the drift test for Part 1 indicated that there was no systematic difference between scores that had been assigned in the pilot study and scores assigned in the drift test: A 95% confidence-interval showed that the difference between the original score and re-score was probably between -.06 standard deviations and +.05 standard deviations. Therefore, it was decided to utilize scores for Part 1 of the 1997 tests that had been computed during the pilot study, and scores for Part 1 of the 2000 and 2001 tests that were computed during 2001.

Agreement between the raters who scored Parts 1 and 2 was remarkably high. Recall that one pair of raters scored Part 1, questions 1 and 2; one pair of raters scored Part 1, question 3; two pairs of raters scored Part 1, question 4, and one pair of raters scored Part 2. Among all these pairs of raters, the correlation between a student's score assigned by the first rater and that same student's score assigned by the second rater ranged from a low of .991 for the "least agreeing" pair of raters to a high of .998 for the "most agreeing" pair of raters. This high agreement indicates that procedures for scoring Parts 1 and 2 erred on the side of caution. The combination of rubrics, anchor items, practice papers and training yielded scoring of very high reliability.

For Parts 1 and 2, each pair of raters scored between four and nineteen separate items; the correlation between the raters was based on the mean of all the items that pair scored. In contrast, raters for Part 3 scored only one item; moreover, the item they scored was particularly involved and difficult to score. The relative difficulty of getting a reliable score for Part 3 was the reason for using the consensus among three scorers in

order to determine students' scores. Given the difficulty of the task, agreement among the raters was reasonably high. The correlation between scores assigned by Rater 1 and those assigned by Rater 2 was .892; the correlation between scores assigned by Rater 2 and those assigned by Rater 3 was .849; the correlation between scores assigned by Rater 1 and those assigned by Rater 3 was .904. The correlation with the consensus score was: for Rater 1, .959; for Rater 2, .909; for Rater 3, .909.

Student Test Scores for Grade 6

Since the early 1980s, the school district where Suburban High School is located has conducted yearly testing using an exam designed by the Educational Records Bureau. Most years, the testing was conducted for all grades from 3 through 10. Since 1996, testing reports have been available on computer disks. The school system has kept an archive containing hard copies of student scores prior to that time. As explained below, Grade 6 test scores from the Traditional and from the two Reform cohorts were used as a covariate in this study.

Transcripts

Suburban High School maintains student transcripts on an automated database. This study used the data base to examine complete transcripts for students in the graduating classes of 1995 through 2001, plus incomplete transcripts available as of spring, 2001 for graduating classes of 2002, 2003, and 2004.

Documents

This study examined the following documents:

1. Syllabi for courses taken by students in the Traditional cohort and in the Experimental cohort;

2. "Complementary Materials" designed as supplementary mathematics resources by Suburban High School teachers;
3. Yearly *School Profiles* published the district in which Suburban High School resides;
4. Yearly *Testing Reports* published the district in which Suburban High School resides.

The Complementary Materials contain page references for readings and problems in the traditional Algebra and Geometry texts that complement topics covered in the IMP modules. Although the Suburban High School did not develop the Complementary Materials until the third year they were utilizing IMP, today every student has access to these traditional textbooks, and their teachers use these Complementary Materials to devise supplemental assignments. The yearly *School Profiles* describe student achievement the preceding year on various measures including participation rate and grades in Advanced Placement exams. *School Profiles* from 1995 through 2001 were available. The yearly *Testing Reports* describe results of Grade 3-10 testing using a test published by the Educational Records Bureau.

Key Informants

This study was completed in close collaboration with two key informants:

1. Mrs. Sullivan, the former mathematics department chair at Suburban High School, who was responsible for implementing the IMP curriculum, and
2. One of the mathematics teachers who first taught IMP at Suburban High School. She spent a year on sabbatical working as an IMP trainer with teachers at Suburban and other high schools, and has since returned to her

teaching position at Suburban High School.

Other information was provided by the current mathematics department chair at Suburban High School.

Data Analysis: Eleventh Grade Algebra Tests

The Algebra Achievement test was intended to measure whether eleventh graders in the Reform cohorts differed from eleventh graders in the Traditional cohort in their understanding of algebra. To address this issue, the primary independent variable analyzed was TREATMENT. It could take on two possible values: 0 for the Traditional (1997) cohort, and 1 for the Reform (2000 or 2001) cohorts.

Covariate: Sixth Grade Test Scores

Beginning in the mid-1980s, students in the school district in which Suburban High School is located began taking a norm-referenced test called the "Comprehensive Testing Program (CTP)" published by the Educational Records Bureau. Until 1993 students completed basically similar tests in the spring of each year, from Grades three through ten. Students received scores in subjects across the curriculum, including mathematics computation, mathematics concepts, and general quantitative ability. In the spring of 1993, when students in the Traditional cohort were in seventh grade and students in the First and Second Reform cohorts were in fourth and third grade respectively, the Educational Records Bureau replaced the CTP II with the CTP III. The "quantitative ability" subtest was retained, but "mathematics concepts" and "mathematics computation" were combined into a single "mathematics" subtest. The Educational Review Board did not create an equated scale that could be used to translate the CTP III scale scores into CTP II scale scores.

While no scale-score equating was done, the Educational Records Bureau did rank student scale scores based on a national norm, with the norm recomputed yearly. As shown in Table 3, 1993 test scores in Suburban High School's district dropped in almost all grades, with a precipitous drop in some grades. The school district's *Testing Report* for 1993 notes that the drop may have been caused by the change from a test that had

Table 3. Yearly Median National Percentile Rank on Educational Records Bureau CTP Quantitative Ability Test

	Year of Testing					
Grade	1991	1992	1993	1994	1995	1996
3	82nd	88th	71st	67th	*	*
4	82nd	86th	82nd	86th	81st	*
5	82nd	88th	88th	91st	91st	88th
6	76th	87th	75th	87th	86th	90th
7	83rd	84th	82nd	85th	84th	88th
8	*	87th	54th	81st	88th	85th

* Data not available for this study

been used each year for the past nine years, to a new test adopted that year. For this reason, it is unlikely that the 1993 scores can be used as a valid control for analyses in this paper.

CTP III scores from the eighth grade, either the spring of 1994 for students in the Traditional cohort or the spring of 1997 and the spring of 1998 for the Reform cohorts, would be a good candidate to use as a control variable, even though there is some indication from median percentile ranks reported in Table 3 that scores in 1994 might still have been lower than in other years when students had more experience with the

particular type of test being used. Unfortunately, a number of student tests taken in 1998 were destroyed by a burst pipe before they could be graded. The destroyed tests included the eighth-grade tests taken by students in the Second Reform cohort.

Since seventh-grade scale scores from the Traditional cohort were probably invalid, and eighth-grade scale scores from one of the Reform cohorts were unavailable, this study used sixth grade-scores from the CTP II (level 4) as completed by students in the Traditional cohort and from CTP III (level E) as completed by students in the Reform cohorts. Four sixth-grade scores were found to be significant predictors of individual students' scores on the Test 1 and Test 2, the parts of the Algebra Achievement test completed by individual students. These four measures were Quantitative Ability, Reading Comprehension, Writing Mechanics, and Verbal Ability.

Test 3 was the portion of the Algebra Achievement test completed by students working in pairs. For each sixth grade measure available as a covariate, three alternate methods of describing the pair score were considered: the mean score for the pair of students completing the test, the maximum score of the pair taking the test (that is, the score of the more able student), and two scores consisting of the maximum and minimum score of the pair of students taking the test. Of the scores available, the best predictor was selected on the basis of the covariate or combination of covariates with the highest adjusted *R-square*. On this basis, the mean quantitative ability score for the two students taking the test was selected. After controlling for mean pair score on Quantitative Ability, none of the other covariates available were statistically significant, so they were not used in the final model.

Although not reported in Chapter 4, the analyses of Test 3 were run using the

alternate choice for covariate of the maximum quantitative ability score from the pair of students completing the test. The results of the alternate analysis were nearly identical to those reported in Chapter 4.

Scale scores on the CTP II taken in sixth-grade by students in the Traditional cohort have not been equated to scale scores on the CTP III taken in sixth grade by students in the two Reform cohorts. Nonetheless, both sets of scale scores are referenced to a "national percentile rank." To control for prior ability, this study has matched sixth-grade scores based on national percentile rank. The legitimacy of this procedure depends on the assumption that a percentile rank in 1992, when students in the Traditional cohort took the CTP II, is comparable to the same percentile rank in 1995 or 1996, when students in the Reform cohorts completed the CTP III. That is, the assumption is that nationwide there was no large change in sixth-grade mathematics competency between 1992 and 1996. This assumption may be questioned: On average, national scale scores in mathematics on the National Assessment of Educational Progress increased by four points between 1992 and 1996 in both Grade 4 and Grade 8 (National Center for Education Statistics, 1997). For this reason, the Analysis section below supplements discussion of results when using sixth-grade test scores as a control by also reporting results without using sixth-grade test scores as a control.

Although sixth-grade scores were equated for this study by using national percentile rank, an unconverted percentile rank is not the best variable to use. In particular, analysis of Part 3, which was taken by students working in pairs, required computing the mean ability of a pair of students. Percentile rank is not an interval scale: for example, it takes a greater increase in ability to move from the 90^{th} to the 95^{th}

percentile than it does to move from the 50th to the 55th percentile. For this reason, sixth-grade scores were converted to z-scores before being used as a covariate. A z-score is the number of standard deviations a particular score is above or below the mean score; assuming ability is normally distributed, every percentile rank can be translated to a particular z-score. The z-scores were created in two steps: first, each national percentile rank was converted to a "national z-score", defined as the z-score that would achieve that rank, assuming normal data. Then, in order to center the mean at zero for the Suburban High School data set, the "national z-scores" were reconverted to "Suburban High School z-scores" by subtracting the mean "national z-score" for all students used in this analysis, and dividing by the standard deviation.

Dependent Variables

The designers of the Algebra Achievement test used it to analyze results for three subscales (Huntley, et al., 2000):

1. "Applied Algebra Problems With Use of Calculators" consisting of all items on each of four forms they designed for Part 1;
2. "Algebra Symbol Manipulation Without Use of Calculators", consisting of all items on each of two forms they designed for Part 2; and
3. "Open-Ended Algebra Problems With Use of Calculators", consisting of three forms they designed for Part 3, each of which contained a single extended problem.

Subscales used for this study are necessarily somewhat different from those used by the Core-Plus authors, because the Core-Plus subscales used items from several forms for each part of the test, whereas this study utilized only one form for each of the three

parts of the test. Also, the pilot study indicated that one particular item on Part 1, Form C fit better on the subscale composed of Part 2 items than it did on the subscale composed of other Part 1 items, and one item on Part 2, Form A fit better on the subscale composed of Part 1 items than it did on the subscale composed of other Part 2 items.

Problem 1.2 on Part 1, Form C, asked students to write an equation for a line, given a graph of that line. The skill required was nearly identical to that tested by problem 14 on Part 2, Form A. In the pilot study for this proposed research, student scores on problem 1.2 correlated more highly with scores on Part 2 than with scores on other items in Part 1, and more highly correlated with problem 14 than with scores on any other item.

Problem 5 on Part 2, Form A asked students to identify an equation describing the relationship between the length and width of a rectangle, given that the length was four meters greater than the width. Except for the multiple-choice format of the question, the skill was similar to that required by items on Part 1 of the assessment that required students to formulate equations to describe algebraic situations. In the pilot study, student scores on problem 5 correlated more highly with scores on Part 1 than with scores on other items in Part 2.

Thus, the present study performed statistical analyses of the following three dependent variables:

Variable 1: Achievement on applied algebra problems in context, as measured by all items Part 1, form C except item 1.2, plus problem 5 form Part 2, Form A (hereafter referred to as Test 1). ;

Variable 2: Achievement on algebra symbol manipulation without context, as

measured by all items on Part 2, Form A except item 5, plus item 1.2 from Part 1, Form C (hereafter referred to as Test 2). ; and

Variable 3: Cooperative solution to an extended open-ended algebra problem, as measured by scores on the single extensive item in Part 3, Form A (hereafter referred to as Test 3).

Test 1, Test 2, and Test 3 were used to address, respectively, the first, second, and third research question posed in Chapter 1 of this study, namely:

i. How do students enrolled in a reform-based curriculum and a semestered block schedule compare to students enrolled in a traditional curriculum and traditional schedule in their ability to solve algebraic symbol manipulation problems? Do the results of this comparison differ depending on students' prior ability?

ii. How do students enrolled in a reform-based curriculum and a semestered block schedule compare to students enrolled in a traditional curriculum and traditional schedule in their ability to interpret and solve challenging algebra problems presented in context? Do the results of this comparison differ depending on students' prior ability?

iii. How do students enrolled in a reform-based curriculum and a semestered block schedule compare to students enrolled in a traditional curriculum and traditional schedule in their ability to collaboratively solve and communicate their solution to a complex open-ended algebra problem? Do the results of this comparison depending on students' prior ability?

Test reliability. For students participating in this study, Test 1 had a reliability (Cronbach's alpha) of .90, while Test 2 had a reliability of .89. Reliability for Test 3 was not computed, since Test 3 consisted of student scores on a single item.

Statistical Methodology: Rules for Establishing Confidence Intervals

Since this study examined three dependent variables, maintaining an experiment-wise error rate of 5 % requires a Bonferroni adjustment, assigning a Type I error rate to each of the three dependent variables of .05/3=.0167.

This study deals with the issue of statistical significance as follows. First, each variable is tested for a significant interaction with prior ability. For Test 1 and Test 2, prior ability is defined as the first principal component of the four grade-six ability scores, because this principal component correlated more highly with both Test 1 and Test 2 than does the quantitative ability score by itself. For Test 3, prior ability is defined as the mean quantitative ability of the two students who took the test.

If a Treatment-by-Ability interaction is not deemed significant, then Treatment alone is tested against each dependent variable, after controlling for prior ability. A 98.33% confidence interval is constructed for each the three effects (i.e., a 98.33% confidence interval around how much the Reform cohorts differed from the Traditional cohort on Test 1, on Test 2 and on Test 3). Thus, there is only a 1.67% chance that the true effect is outside the confidence interval, and a 95% probability that all three effects are actually within the reported confidence interval.

A 95% confidence interval is also reported for each of the three effects. For each test, there is only a 5% chance that the true effect is outside this confidence interval; overall, there could be as little as an 85% probability that all three effects are actually

within the reported confidence interval.

Statistical Models Used

The first two dependent variables (Algebra Problems in Context and Symbol Manipulation) consist of a student's average score on a large number of items and can be assumed to be on approximately a ratio scale. The pilot study found that residual scores on these variables, after entering controls, were approximately normally distributed. For ratio-scale data with normally distributed residuals, statistical methods based on the General Linear Model have optimum power and appropriate error rates. For this reason, when controlling for covariates the analysis of the first two dependent variables was performed using Ordinary Least Squares (OLS) Linear Regression, which is based on the General Linear Model. This methodology yields results identical to what would be reported by an Analysis of Covariance, or ANCOVA, but has the advantage of yielding effect sizes that can be interpreted. When not controlling for covariates, the analysis of the first two dependent variables was performed using an independent-samples t-test, which is mathematically equivalent to OLS regression.

The third dependent variable (Cooperative Solution to an Extended Open-Ended Algebra Problem) uses an ordinal scale of student scores, taking on possible values of 0 to 4. Analysis for this variable was identical to that utilized for the first two, except that Ordinal Regression was used instead of Ordinary Least Squares Regression. There are two varieties of Ordinal Regression that are commonly used with such data, Probit Regression and Logistic Regression. Unless the student ability distribution is very unusual, both types of regression will provide nearly identical results in terms of p-values, but the interpretation is slightly different. Each highlights a different and

important aspect of the data. Probit analysis provides an effect size that can readily be compared to output from the analyses of Test 1 and Test 2. Logistic regression analysis provides an odds ratio that is more easily related to student responses that were actually observed. In the interest of clearly explaining the observed results, both types of analysis are reported below.

Supplemental Analyses of Specific Skills

In addition to the omnibus statistical tests for differences in student achievement on Test 1, Test 2, and Test 3 a number of supplemental analyses are reported in this study. These analyses provided a finer-grained picture of how the Reform cohorts differed from the Traditional cohort on specific algebra skills contained within Test 1 and Test 2.

To facilitate the finer-grained analysis, two sub-scales of items were formed to examine the following specific skills:

1. Skill 1: Formulating Mathematical Models (Part 1, problems 1.1, 1.3, 1.5a, 1.5b, and Part 2, problem 5). Reliability as measured by Cronbach's alpha: .70
2. Skill 2: Interpreting Algebraic Models (Part 1, problems 4.1,4.2,4.3,4.4, and 4.5). Reliability as measured by Cronbach's alpha: .89

Differences between the Reform cohorts and the Traditional cohort on these two subscales were investigated using an independent-samples *t*-test. Items from Test 1 and Test 2 that were not on the specific subscales were examined individually. For dichotomous individual items, that is, items that were scored as right/wrong, a Pearson

Chi-square statistic was computed from a cross-tabulation table. On items for which students could receive partial credit, a Wald Chi-square statistic was computed from an ordinal Logistic Regression Analysis. The Logistic Regression Analysis is a generalization of the cross-tabulation method used to examine dichotomous items. Altogether, the supplemental analysis of specific skills compared the Reform cohorts to the Traditional cohort on 25 measures. To guard against over-interpreting results that occurred by chance, the Reform cohorts were deemed to be different from the Traditional cohort on one of the twenty-four measures if the statistical significance level for that measure reached a Bonferroni-adjusted .05/25 = .002 level.

Data Analysis: Student Transcripts

Because of the change to a semestered block schedule, no student at Suburban High School can complete mathematics courses entailing precisely the same number of hours as offered under the prior scheduling system. The semestered block schedule makes more courses available to each student, with each course lasting fewer hours. Students who allocate the same number of courses to mathematics as they would have done under the prior schedule will actually complete fewer hours of mathematics instruction. Students who enroll in more mathematics courses through semestered block scheduling than they would have under the prior schedule actually do receive more hours of mathematics instruction.

The semestered schedule affords students the opportunity to study more mathematics, if they are motivated to do so. In an extreme case, a student could hypothetically complete eight sequential mathematics courses over her/his high school career. Mathematics teachers believed that having students begin their study of high

school mathematics with the IMP curriculum would provide the motivation to study more mathematics. This belief is consistent with research to date. Webb (in press) investigated three schools in California that were the first to offer at least three years of IMP. He found that 64% of students at these schools who started IMP in Grade 9 actually completed at least four years of high school mathematics. In comparison, 38% of students at these schools who started a traditional sequence with Algebra I in Grade 9 actually completed at least four years of high school mathematics. This difference in course taking was significant at the .01 level.

The question remains: Did students really register in additional or more advanced mathematics under the IMP curriculum and semestered block schedule at Suburban High School than under the prior curriculum and schedule? The Secondary Research Questions listed previously break this larger question into the following more detailed questions:

iv. How did students enrolled in a reform-based curriculum and a semestered block schedule differ from students enrolled in a traditional curriculum and traditional schedule in the number of registered mathematics class hours by the end of Grade 12?

v. How did students enrolled in a reform-based curriculum and a semestered block schedule differ from students enrolled in a traditional curriculum and traditional schedule in participation in advanced courses, as measured by the number of registered hours in advanced mathematics classes by the end of Grade 12, by the number of students enrolling in Advanced Placement courses, and by scores on Advanced Placement tests?

This aspect of the study is addressed using two sources of data: student transcripts, and annual School Profiles published by Suburban High School.

The transcripts were used to determine course enrollment. For each student each year, the hours registered in mathematics class were computed by multiplying the number of mathematics courses in which the student was enrolled times the number of days per course times the number of hours per course per day

The number of hours registered in advanced mathematics classes was computed the same way, with an advanced class being defined as a class taken after the core requirements were completed. For the traditional curriculum, advanced classes were defined as those courses taken after successfully completing Algebra 2. For the reform curriculum, advanced classes were those courses taken after successfully completing Integrated Math 4.

Student transcripts were also used to determine the number of students enrolled in Advanced Placement courses each year, by grade level. Students completing Advanced Placement courses had the opportunity to earn college credit in mathematics by taking an Advanced Placement exam administered yearly by the College Board, and many students did so. This study obtained student grades on Advanced Placement exams from annual School Profiles published by Suburban High School.

Groups to be Compared

Students in the Traditional cohort who were tested in the spring of 1997 had used a traditional curriculum and schedule from ninth grade through the time of testing, when they were in eleventh grade. However, when those same students were in twelfth grade during the 1997-98 school year, Suburban High School adopted a pilot block schedule at

all grade levels. Since they were *not* enrolled in a traditional schedule in twelfth grade, it would not be appropriate to use these students' transcripts to address research questions iv. and v. Instead, this study used transcript data from the two preceding cohorts at Suburban High School: those students who graduated in the spring of 1996, and those students who graduated in the spring of 1997. Both of these cohorts of students were enrolled in a traditional schedule and curriculum throughout their high school tenure.

Complete transcript data was also available for the First Reform cohort, who graduated in the spring of 2001. As of this writing, students in the Second Reform cohort have not yet completed high school. Therefore, the transcript analysis in this report compared the registered hours of mathematics studied for students in the First Reform cohort, to the registered hours of mathematics for students in the Traditional cohorts who graduated in the spring of 1996 or the spring of 1997.

As noted previously, student transcripts were also used to identify the number of students enrolled in Advanced Placement mathematics courses each year by grade level and School Profiles were used to obtain student grades on Advanced Placement exams. Data on Advanced Placement enrollment was available each year from 1990-91 through 2000-01 and is reported. Data on Advanced Placement exam scores was available each year from 1994-95 through 2000-01 and is reported.

Statistical Methodology

Formal hypothesis testing could obscure important information about differences in the amount of mathematics that the Reform and Traditional students studied. This is because significance tests ask the question: "Is there a non-chance difference between the groups on some particular population parameter?" The parameter might for example

be an adjusted or unadjusted mean value (as reflected in t-tests or linear regression), or a median value (as in "nonparametric" tests like the Mann-Whitney U), or the odds of attaining a certain goal (as in Logistic Regression Analysis). Such population parameters report overall summaries for a set of data. They are likely to obscure possibly interesting patterns of differences between cohorts in detailed course-taking patterns. For example, reporting the mean number of hours enrolled in advanced mathematics courses might miss a pattern in which the bulk of students in a particular cohort enrolled in fewer hours, while a few students in that same cohort enrolled in substantially more hours. Other summary statistics have similar limitations.

Because of the limitations of traditional hypothesis testing, the analysis of the course registration data will consist of inspection of graphs depicting the registered mathematics hours for each group. In some cases, apparent differences between groups are tested for "statistical significance" to determine if they are likely to have occurred by chance. These post-hoc significance tests, which were conducted after identification of patterns from the graphs, should be viewed as exploratory analyses, not confirmatory hypothesis tests.

After examining differences between the AP exam scores of students who had used the semestered block schedule and the IMP curriculum and the AP exam scores of students in earlier years, a post-hoc Logistic Regression Analysis was done to see if the apparent differences between groups in the Advanced Placement Calculus BC exam could easily be dismissed as being chance variation. Because the statistical analysis of Advanced Placement Calculus BC exam grades was conducted after noticing a pattern in the data, it too must be viewed as an exploratory analysis.

Data Analysis: Documents and Key Informants

Research question vi states:

vi. According to administrators and faculty who assumed critical responsibilities for implementing the shift to a semestered block schedule and reform-based mathematics curriculum, what school administrative policies and what unique aspects of the school and community affected mathematics course enrollment and mathematics achievement under the new schedule and curriculum?

Data collected from policy documents and key informants was sought in order to describe the key administrative and curricular decisions, as well as other variables that may have contributed to differences between the Traditional and Experimental cohorts in mathematics course-taking or in algebra achievement. Information gathered about these topics was used to inform the discussion in Chapter 5 of this study.

In addition to providing information to address the sixth research question, key informants also provided details about specific course content. This information was used to analyze student opportunity to learn specific topics of interest identified in the fine-grained analysis of specific skills within the Algebra Achievement test.

Summary of Method

This study compared the algebra achievement and mathematics course-taking patterns of two groups of students at a suburban high school in the eastern United States. The Traditional cohort used a traditional curriculum and schedule throughout high school. The Reform cohorts attended the same high school with primarily the same

teachers as the Traditional cohort, but they used a semestered block schedule and the IMP (Fendel, et al., 1997) curriculum.

The analysis of algebra achievement compared students on three subscales: "Algebra in Context," "Symbol Manipulation," and "In-depth Algebra Problem Solving." These subscales were derived from an algebra assessment designed by the Core-Plus Mathematics Project (Huntley, et. al, 2000), and were administered to students in each cohort at the end of Grade 11. Items on the assessment were scored by a panel of experts using a scoring rubric, anchor items, and practice papers. Grade 6 test scores were used as covariates to control for differences between the two cohorts in initial ability. The analysis also investigated whether there is any interaction between treatment and prior ability as measured by sixth grade test scores.

In order to analyze mathematics course-taking student transcripts were reviewed to compare the number of hours that students from each cohort spent registered in mathematics classes in Grades 9-12. Hours registered in all mathematics classes, and hours registered in advanced mathematics classes were compared separately. Instead of formal hypothesis testing, a graphical display was used to detect important differences between the two groups in mathematics course-taking.

Yearly student enrollment in Advanced Placement courses was reported, using data obtained from student transcripts. Student grades on Advanced Placement exams administered by the College Board were obtained from annual School Profiles published by Suburban High School.