

DRAFT

***A Tale of Synergy: The Joint Impact of 4x4 Block Scheduling and an NCTM
Standards-based Curriculum on High School Mathematics Achievement***

***By Dr. Steven Kramer, University of Maryland
Regina Keller, Mathematics Curriculum Specialist
April 1, 2003***

Abstract

This study investigated mathematics learning at a suburban United States high school that simultaneously adopted a semestered (4 x 4) block schedule and the Interactive Mathematics Program (IMP), a curriculum designed to implement the National Council of Teachers of Mathematics (NCTM) *Standards*. Previous research has often found that at sites where a block schedule was adopted without changes to mathematics curriculum and instruction, mathematics achievement has declined. In contrast, when a semestered block schedule and the IMP curriculum were implemented jointly, with extra time allocated to planning and staff development, the two innovations were followed by improvements in student mathematics achievement.

Statement of the Problem

This study investigated the joint effects of two reforms that are beginning to have widespread impact on mathematics instruction in high schools throughout the United States and Canada. The first reform is adoption of a problem-centered mathematics curriculum designed to conform to the National Council of Teachers Mathematics' (NCTM's) *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). The second reform is adoption of a semestered block schedule.

Standards-based Curricula.

In 1989, NCTM promulgated the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). The *Standards* called for mathematics classes to include less traditional teacher exposition and student practice, while increasing the amount of project work, group and individual assignments, and discussions between teacher and students and among students (NCTM, 1989, p. 10). The *Standards* reemphasized and strongly endorsed earlier statements by the NCTM that “problem solving must be the focus of school mathematics” (NCTM, 1980, p. 2; NCTM, 1989, p. 6). Later *Standards* documents (NCTM, 1991, 1993, 2000) provided a vision that was consistent with a problem-centered approach to mathematics instruction.

The Interactive Mathematics Program (IMP) was one of several high school mathematics curricula developed during the 1990s to conform to the vision described in the NCTM *Standards*. There has been considerable controversy about the worth of these new curricula (see, for example, “An Open Letter to United States Secretary of Education, Richard Riley”, 1999). In a well publicized critique of the IMP curriculum in particular, a Berkeley University mathematics professor noted that after reading the curriculum he felt IMP would be appropriate

for the majority of students, but it would not provide an adequate technical foundation for students who wished to study advanced mathematics (Wu, 2000).

Despite these criticisms, research to date has generally shown that IMP and other *Standards* based curricula tend to have positive effects on student achievement (Koedinger, Anderson, Hadley, & Mark, 1997; Huntley, Rasmussen, Villarubi, Sangtong, & Fey, 2000; Merlino, Wolff, & Tolbert, 2000; Webb, in press). Nonetheless, the evidence is not conclusive, as these studies have generally been small in scale, with the number of schools in each study ranging from one to six. Further, in all cases, teachers using the new curricula have been volunteers, and in many of the studies the students have been volunteers as well. More research is needed to determine whether earlier results can be replicated, to extend study of reform mathematics to situations where results cannot be attributed to a volunteer effect, and to address the ability of reform curricula to prepare students for studying advanced mathematics.

Block Scheduling

Over the last decade, block schedules have become increasingly popular in high schools throughout the United States and Canada. Under a traditional high school schedule, each student typically enrolls in seven courses at a time. Each course meets daily for about 40-50 minutes over a 180-day academic year. In contrast, under the most common type of block schedule, the 4x4 or “semestered” plan, each student typically completes four courses in each semester. Each course meets daily for about 80-90 minutes over a 90-day semester.

Block scheduling has become popular among administrators for a number of reasons. School atmosphere tends to be calmer, and many schools have reported fewer discipline problems under a block schedule (Carroll, 1994a; Hackman, 1995; Meadows, 1995; Reid, 1995a, 1995b; Sessoms, 1995). With fewer classes per day, there are fewer transitions, so students can spend more time in class instead of walking through hallways between classes. Moreover, some

Draft

surveys indicate that both students and parents tend to prefer a semestered block schedule to a traditional schedule (Stevens, 1976).

Despite the advantages cited above, there is reason to be concerned that changing to a block schedule can actually decrease student achievement, especially in mathematics. Such a decrease was noted in Canada, where block schedules have been used since the 1970s, well before they became popular in the United States. Several large-scale Canadian studies have reported that mathematics achievement under a semestered block schedule was lower than under a traditional schedule (Raphael, Wahlstrom, & McLean, 1986; Marshall, Taylor, Bateson, & Brigden, 1995; Wild, 1998).

In contrast, anecdotal reports (Kramer, 1996), as well as some achievement data from North Carolina (Averett, 1994; North Carolina Department of Public Instruction, 1997, Zhang, 2001) indicate that mathematics achievement may improve if a semestered block schedule is adopted at a school that implements instructional changes consistent with those called for by the NCTM *Standards* (NCTM, 1989). However, this possibility has never been systematically studied.

It is also important to investigate how a block schedule affects mathematics achievement at a site where course syllabi have been adjusted to fit the schedule, by spreading mathematics content over a greater number of courses. In the United States, students generally take more courses per year under a block schedule, with fewer total minutes allocated to each course. Thus, less content can be covered per course but more courses are available. One solution may be to change syllabi when moving to a block schedule, so that the same content is covered over a greater number of courses (Harter, 1994; Kramer, 1996). Unfortunately, in most cases, research on the achievement effects of block scheduling has been conducted at sites where this approach

has not been implemented¹. The current study investigated achievement at a school that did implement this approach, replacing a three-course core sequence (for most students, algebra 1, geometry, and algebra 2) with a four-course core sequence utilizing the IMP texts.

Research Opportunity at Suburban High School²

We began collecting data for this research project in the spring of the 1996-97 school year. At that time, Regina Keller, the second author of this paper, was chair of the mathematics department at Suburban High School and mathematics supervisor for Suburban School District near Philadelphia, Pennsylvania. Suburban is the only high school in the school district. The immediate area is rich in institutions of higher education. While students at the school have a wide variety of backgrounds, average achievement on standardized tests has traditionally been high.

During the 1996-97 school year, Suburban High School initiated a semestered block schedule for all ninth graders. In that same year, all ninth graders except honors students who had already completed Algebra 1 were enrolled in IMP, as a phase-in of this new *Standards*-based curriculum. These students and their successors continued in this schedule/curriculum throughout high school. Beginning in 1997/98, IMP was utilized as the core curriculum for all students including those in the Honors program. In addition to implementing a new schedule and curriculum, Suburban High School allocated significant resources to planning and staff development. This paralleled the conditions previous qualitative research indicated might improve mathematics achievement under a semestered block schedule (Kramer, 1996).

The primary focus of this study is on a quantitative analysis of changes in student learning at Suburban High School after the semestered block schedule and IMP curriculum

²The one exception we are aware of is a study by Zhang (2001). Zhang reported that North Carolina schools adopting a semestered block schedule frequently allowed lower-achieving students to study Algebra 1 as a sequence of two one-semester courses instead of the traditional year-long course. It is noteworthy that algebra achievement improved in North Carolina schools adopting a semestered block schedule—the only large-scale example we know of which documents improved mathematics achievement under a block schedule.

² “Suburban” is a pseudonym.

were adopted. However, the quantitative results will be misleading unless readers can develop a good qualitative understanding of what was done to achieve those results. To this end, we turn to a first person retrospective account of how the new program was implemented, prepared by the second author of this paper, who was instrumental in the implementation. This account is contained in the next two sections: “History of the Implementation” and “How did the Classroom Change?” To check the validity of Ms. Keller’s account, we asked mathematics teachers who had participated in implementing the new program and schedule to review what Ms. Keller had written. We also reviewed student transcripts to confirm information in the account.

Implementation of the new high school curriculum and schedule began in the fall of 1996, but preparation for the change really began in the spring of 1991, with the adoption of a new Middle School mathematics curriculum for the 1991/92 school year. At that time, students who were later members of the Traditional comparison group we tested in 1997 and 1998 were just about to enter sixth grade. Ms. Keller’s account begins from there.

History of the Implementation

In the spring of 1991, as Mathematics Supervisor in Suburban School District, I was given the challenge of selecting a curriculum to serve the newly merged middle schools. The steering committee that facilitated the combining of the two middle schools also determined that all classes would be heterogeneously grouped. I was immediately concerned. How could the traditional texts, which stressed arithmetic operations for three years, be successful for such a diverse group of students? After I spent many unsuccessful hours of research, my Assistant Superintendent handed me materials from a district in Vermont that looked promising.

The program was called “Visual Mathematics” and was developed at Portland State University in the Math Learning Center (Foreman and Bennet, 1991). It was a constructivist

Draft

approach (though “constructivist” was not yet in my vocabulary). What I recognized was that it had a scope and sequence that encompassed more concepts in mathematics than simple arithmetic, was centered on open-ended problem solving, used cooperative groups, emphasized writing along with other communication skills, and stressed understanding of the mathematics, not just memorization of algorithms. I was sold.

The veteran sixth grade teachers who piloted the program in the fall were enthusiastic and campaigned to use the material for sixth, seventh, and eighth grades. At that point it became obvious that the staff and I needed more training to deliver the curriculum effectively. The summer of 1992 started an annual series of 30-hour professional development summer experiences. I decided that all teachers in Grades 5-12 should be trained in the middle school program. I felt it was important that 5th grade teachers knew the type of mathematics their students would be facing in the new middle school and the secondary teachers needed to know what mathematics had been taught and how it was delivered.

The staff from the Math Learning Center delivered excellent professional development to all of us each summer for the next three years, 1992-94. We discovered how students learn mathematics and began to change our pedagogy. No longer were we the “sage on the stage” but we learned to be facilitators of learning. When the first group of students completed the program in 1994, their standardized test scores were excellent. Since there had been plenty of opposition to the implementation of such a unique approach to teaching and learning mathematics, I was very pleased. I had taught parent classes, spoken before the board of school directors on numerous occasions, and given training to the administrative staff. Even with all this effort, the program would not have succeeded without the strong support of the Mathematics Department and the top administrators.

Suburban School District was very progressive. In 1995, the district decided to
Draft

investigate alternative methods of scheduling. I was selected to be on the task force that spearheaded this effort. The task force reviewed and observed many types of schedules during the next two years and finally approved a semestered block schedule. The new schedule was to be piloted with the ninth graders in the 1996-97 school year. It was to be implemented for the entire school in 1997-98.

This process produced angst for the entire school community. Although the high school would continue with homogeneous groups, parents and teachers alike were unconvinced that the students would truly receive the same amount of information. Obviously, 80-minute classroom periods for a semester did not equate to 45-minute classroom periods for the entire year. What about the “gap”? If a student had Spanish 1 in the fall of one year and didn’t receive Spanish 2 until the spring of the following year, would he/she remember anything?

While the task force was deliberating, the Mathematics Department decided to become proactive. If we went with a new schedule, how could we make it work for our students? The first recommendation was to increase the graduation requirement from three year-long mathematics courses to four semester courses. This change would even out the time factor. There was no resistance to this suggestion since all students would be taking eight courses a year instead of the usual six or seven. There was plenty of room for another math selection.

The one concern that had plagued me as the Advanced Placement (AP) Calculus teacher was that I was spending more time reviewing algebraic skills than I was teaching calculus concepts. I had hired most of our teachers and had total faith in their ability to teach these students, and I knew that they followed the curriculum guides we had produced. Could the problem be that algebra concepts were taught in two discrete courses in Grades eight and ten? The Department began the task of integrating algebra concepts throughout the curriculum. The premise was that if the students had algebra in every course instead of in two isolated courses,

Draft

just maybe they would internalize the material. While we were at it, we decided to integrate all the traditional material: Geometry, Trigonometry, Discrete mathematics, Statistics, and Probability. This approach might help the “gap” factor, too.

We began this process with Algebra concepts. We spread them over a four-course sequence allowing them to build each year. We used the NCTM *Standards* (NCTM, 1989) as our guide and had many spirited discussions. Did we really need to spend four months on factoring? What was the purpose of factoring anyway? How many two column proofs were necessary? Was statistics really important and not an optional topic? After a year of difficult departmental work, we came up with a blueprint for our new integrated program.

We definitely needed some validation for our product. Early in 1996, I convened a group of about 15 parents from the district who were either college mathematics educators or mathematicians. The department presented the new program and asked for feedback. The parent group offered a few suggestions for topics that we had underemphasized and were enthusiastic about the concept of integrating all topics. They did point out one potentially fatal flaw. What materials would we use? (When we started this process we did try to look at the “Integrated” texts that were on the market, but they were integrated in name only, so we relied on our own knowledge and experience and decided to cross that bridge when it became necessary.) Luckily, the NCTM conference was in a nearby city that spring so we all attended. The goal was to find curriculum materials that would work with our model.

We found it in a curriculum originally suggested by one of the parents on the review committee. It was designated a reform curriculum sponsored by the National Science Foundation. We didn’t know what a “reform curriculum” was exactly, but we did recognize that it embodied the same philosophy as our Middle School program. The Interactive Mathematics Program’s scope and sequence matched ours almost perfectly. We needed to move a few topics

Draft

around and add some statistics concepts we had missed, but it worked. We did a thorough investigation of this program, including observing classrooms where it was being piloted and taking a course in one of its stand-alone units. I sold it to the Board of School Directors, and we were on our way to phasing in the program one grade each year. That summer the teachers in Grades 6-12 again did a 30-hour training course in IMP.

We did have two challenges in using IMP. First, the program was written for a traditional schedule (approximately 45 minute classes). We handled this by making changes as needed in the way topics were addressed; for example, sometimes homework was done in class and other times classwork was used for homework. We also took advantage of the supplemental problems and used them when needed.

Second, IMP was intended for heterogeneous grouping, but that approach would not fly in the district. (After a couple of years the middle school went back to ability grouping). We needed to take our blueprint and adapt it for our three ability levels of students. We originally had the Honors students finishing four of the five units in each course, the College Prep classes doing three units and the Academic Assisted doing two units each semester. That schedule would allow all students to take more mathematics than they had taken in the traditional schedule. I should note that students were not locked into any specific level. Students moved both up and down from Honors to College Prep and to Academic Assisted.

We began implementing the new curriculum in the fall of 1996. All ninth graders were given IMP Course 1 except those students who were enrolled in Geometry Honors. If reform students elected to take a second course that year, we offered Course 2 the second semester. At the same time, we began using IMP to teach the highly gifted 8th grade students who normally took Algebra 1. We expected them to complete both Course 1 and Course 2 during 8th grade.

Other middle school top students would take Course 1.

The following year we added Courses 3 and 4 at the high school while deliberately phasing out the traditional courses of Geometry, Algebra 2, and Pre-Calculus. One teacher had to double up on her 30-hour training courses over the next two years to be prepared to teach the entire curriculum by the end of the second year. The program's professional development also required an additional 30 hours to be completed during the school year—either after school, on Saturdays, or during in-service days. The department members were very professional in their dedication to this process; they often did not receive any compensation for their time.

We continued this process for the next three years: summer training, school year training and revision of curriculum. As we became more experienced, we managed to add an additional unit to each of the three levels of IMP courses: Honors, College Prep and Academic Assisted. I again experienced opposition from a conservative group of parents and school directors. With the department's help, we kept the program. It helped that our school's scores on the State Assessment for the first cohort (class of 2001) were the highest in the state for a comprehensive high school.

Under the semestered block schedule, we needed to add some courses to our sequence since some students could be finished with the core curriculum at the end of ninth grade. Parents made it clear that they wanted the students to have as many mathematics courses as possible. This was a great opportunity for us. We broke Pre-Calculus into two courses: Functional Analysis and Discrete Analysis, each with Honors and College Preparatory sections. The new Functional Analysis course was viewed as a prerequisite to calculus, while the Discrete Analysis course was an advanced elective available to interested students. Previously, advanced students had a choice of two year-long AP Calculus courses, either "Calculus AB" or "Calculus BC". We replaced these year-long courses with a sequence of two semester-long courses: the first covered

Draft

Calculus AB; and the second course, Calculus BC, took up where the first left off. We also offered a College Prep Statistics course and AP Statistics. We continued to offer a course entitled “Contemporary Mathematics,” intended for students who may have difficulty with Integrated Math 4. (This course ended up being taken by about 5% of the students in each graduating class.) In addition, some students enrolled in an “SAT prep” class, which carried one-half credit of mathematics and one-half credit of English.

Under the semestered block schedule, in any given school year students could potentially “double up” to enroll in two sequential mathematics courses, one each semester. This opportunity made it theoretically possible for Honors students who started with Integrated Math 2 or 3 to finish the IMP sequence and subsequently enroll in up to five additional courses. For example, a student beginning with Integrated Math 3 Honors at the high school could potentially take the following:

Ninth Grade:	Integrated Math 3 Honors and Integrated Math 4 Honors
Tenth Grade:	Functional Analysis Honors and Discrete Analysis Honors
Eleventh Grade:	Calculus AB and BC
Twelfth Grade:	Statistics AP

In theory, a student beginning with Integrated Math 1 who decided to take two semesters of mathematics each year could complete the four IMP courses and enroll in up to four additional courses. College Preparatory courses available included Discrete Analysis, Functional Analysis, Statistics, the half-credit SAT prep course, and (in 2002 only) Introduction to Calculus Honors. After completing the IMP sequence, Academic Assisted students could enroll in Contemporary Mathematics, in a higher-level IMP course that completed the units not included in the Academic Assisted curriculum, or in other College Preparatory courses.

In reality, most Honors students doubled up at least twice, while the academic assisted

Draft

students rarely did. College prep students often doubled up for one year. The department did not recommend doubling up. The material became so rigorous in the last two Integrated Math courses using the IMP texts that we didn't want to rush students into such difficult concepts. However, parents and guidance counselors were committed to having each student take as much mathematics as possible, especially when the program was first instituted.

As we began the process in the 1996-97 school year, I realized that we would need some "proof" that this approach was beneficial for students. Since we were the first suburban school district to institute a total conversion to a reform curriculum and a block schedule, we didn't have any models or data on which to rely. There were no studies for our situation. However, I read an article in the *NASSP Bulletin* (Kramer, 1997b) by Steve Kramer, a doctoral student at the University of Maryland who had studied block schedules with traditional mathematics and found the combination to be lacking. I contacted the author and suggested that he do a longitudinal study at Suburban High. I was convinced that the department had implemented the block schedule and a reform curriculum successfully. Steve and I worked closely together, beginning data collection in the spring of 1997.

How did the Classroom Change?

Prior to changing the high school schedule and curriculum, the normal sequence of a 45-minute mathematics period had been for the teacher first to go over the homework, then to present new material, and finally to give the new homework. With a block schedule and an IMP curriculum this changed. Teachers agreed that just giving lectures for 80 minutes would make a very boring learning environment. They welcomed the activities that IMP offered to break up the time and have the students become more actively involved. Students could now review their homework in groups, and the teacher would need only to review problems with which the entire group or class had difficulty. The teachers spent less time being the "sage on the stage" and had

Draft

the students more actively involved with the lesson. The students used manipulatives and went through a discovery process to comprehend the new concepts.

All teachers and students became comfortable with the use of the overhead projector. Teachers began moving around the room more as students worked in groups on some concept. Teachers got off the stage completely. They learned to ask probing questions instead of giving quick answers.

Previous to adoption of the new schedule and curriculum, use of manipulatives and inquiry learning were the exceptions. Further, although many staff members had used cooperative groups and writing tasks for some activities, with the new program these techniques became the norm for the classroom. In addition, students were now expected to make presentations on a regular basis. They learned to attack open-ended problems with some confidence.

These activities required a change in grading. Alternative assessments included Problems of the Week (POWs), Portfolios, and Presentations. Teachers began using rubrics for the first time.

Teachers noted that in their traditional classes, such as Functional Analysis, Calculus, and Discrete Analysis, they also used the discovery method and groups more than they had in the past (before block and IMP). Under the semestered block schedule and IMP, teachers recognized that students became more facile in their ability to communicate and more flexible in their approach to problem solving.

Grade 11 Algebra Achievement Test

Testing the Traditional cohort. In order to assess their understanding of algebra, in the spring of 1997 we administered an Algebra Achievement Test to all eleventh graders at Suburban High School. Eleventh graders tested in 1997 comprise a “Traditional” cohort who

Draft

used a traditional schedule and traditional curriculum throughout Grades nine through eleven.

The Algebra Achievement Test was designed as part of a study investigating the Core Plus Mathematics Project (Huntley, Rasmussen, Villarubi, Sangtong, & Fey, 2000). The test was specifically designed to compare the effects of a Standards-based curriculum to those of more conventional curricula. It is organized into three parts. Part 1 focuses on students' ability to understand and solve algebra problems presented in context, as is typically emphasized by IMP and other reform curricula. Part 2 focuses on context-free symbolic manipulations of a type typically emphasized by traditional algebra texts. Items in Part 2 were adapted from released ACT examinations and from items that commonly appeared on college placement tests. Part 3 requires collaborative work on a single extensive open-ended applied problem and is completed by students working in pairs. The complete Algebra Achievement Test appears in Kramer (2001).

Testing the Reform cohorts. In the spring of 2000 and the spring of 2001, eleventh graders at Suburban High School again completed the Algebra Achievement Test. Eleventh graders tested at these times comprise the "Block Schedule/IMP" cohorts. They used IMP as their core high school curriculum and were enrolled in a semestered block schedule throughout high school.

Grade 12 NAEP blocks.

While algebra has been at the heart of high school mathematics for many years, teachers at Suburban High School were concerned that their new curriculum and schedule might also have important effects on student achievement in Probability/Statistics and in Geometry, areas of mathematics that the Algebra Achievement Test did not address. To assess student achievement in these content areas, we decided to administer the three publicly released blocks of items from the 1996 Twelfth Grade Mathematics Assessment administered by the National Assessment of

Draft

Educational Progress (NAEP). Approximately two-thirds of the items on the NAEP blocks addressed Data Analysis/Probability/Statistics or Geometry/Meaning. The three blocks utilized for this study can be obtained online at <http://nces.ed.gov/nationsreportcard/ITMRLS/pickone.asp>. To view the items, scroll down to the “Search by Block” option, select “Mathematics” and “Grade 12,” click on “Continue,” and then select each the three blocks “1996-12M10,” “1996-12M12,” and “1996-12M13.”

Testing the Traditional cohort. In the spring of 1998, twelfth graders in the Traditional cohort at Suburban High School completed the three NAEP blocks. Unfortunately, we did not obtain the NAEP blocks until the 1997-98 school year, after the entire school had switched to a semestered block schedule. Unlike the Algebra Achievement Test which had been administered to all eleventh graders, the NAEP blocks were administered in mathematics class only to students enrolled in mathematics during the second semester of their senior year. Twelfth graders tested in the spring of 1998 used a traditional curriculum throughout Grades nine through twelve, a traditional schedule in Grades nine through eleven, and a semestered block schedule in Grade 12.

Testing the Pilot cohort. In the spring of 2000, twelfth graders at Suburban High School again completed the three NAEP blocks. As with the Traditional cohort, only students enrolled in mathematics during the second semester of their senior year completed the test. Students tested in the spring of 2000 comprise the Pilot cohort. All students in this cohort utilized a semestered block schedule throughout Grades 9-12. Non-Honors students in this group were the first to utilize IMP as their core curriculum. Importantly, Honors students in the Pilot cohort completed a traditional core mathematics sequence that they had begun with Algebra 1 in the eighth grade.

Contrasting achievement of Honors students from the Pilot cohort to that of Honors
Draft

students in subsequent Block Schedule/IMP cohorts provided particularly useful information. It enabled us to compare achievement of high-ability students who studied a Traditional curriculum taught under a semestered block schedule to that of high-ability students who studied the IMP curriculum taught under a semestered block schedule.

Testing the Reform cohorts. In the spring of 2001 and the spring of 2002, twelfth graders from the Block Schedule/IMP cohorts who were enrolled in mathematics class during the second semester of their senior year again completed the three NAEP blocks. Both Honors and non-Honors students in the Block Schedule/IMP cohorts used IMP as their core high school curriculum and were enrolled in a semestered block schedule throughout high school.

Other Data Analyzed

In addition to Algebra Achievement Test scores and NAEP test scores, students in the Block Schedule/IMP cohorts were compared to students in Traditional and Pilot cohorts using other available data. Additional data included student transcripts and scores on Advanced Placement exams administered by the College Board.

Scoring Procedures

The tests we administered contained many constructed response items. An expert panel comprised primarily of college professors in mathematics scored these items, following professional standards used by NAEP, as described in Bourgeacq, et al. (1997). Members of the expert panel participated in two days of training using the scoring rubric, anchor papers, and practice papers before scoring the Algebra Achievement Test, and an additional day of training before scoring the NAEP blocks.

Three members of the expert panel scored Part 3 of the Algebra Achievement Test, which we viewed as the most difficult item to score. We split the remaining constructed response items into groups, each of which was scored by two members of the expert panel. After scoring, the

Draft

raters met and came to a consensus score on any items where their independently assigned scores disagreed.

For one of the NAEP blocks, we could not compute inter-rater agreement because the raters did not preserve a record of the scores they assigned prior to reaching consensus. On constructed-response items in the other two NAEP blocks, we measured inter-rater agreement using Cohen's Kappa for items with 0/1 scores and using the correlation between rater scores for items with ordinal scores. On all items measured, inter-rater agreement was greater than .90. Kramer (2002) provides a detailed discussion of inter-rater reliability on the Algebra Achievement Test.

Analysis of the Grade 11 Algebra Achievement Test

Methodology

We viewed each of the three parts of the Algebra Achievement Test as an equally important measure of student achievement. Therefore, to maintain an overall 5% Type I error rate for our analysis of the eleventh grade testing, we utilized a Bonferroni adjustment, allocating a confidence level of $.05/3=.017$ for statistical testing within each of Part 1, Part 2 and Part 3.

While Part 1 and Part 2 of the Algebra Achievement Test can each be treated as an individual subscale, we found it more enlightening to analyze nine discrete skills tested by the items in Parts 1 and 2. We divided several of these skills into more than one subscale depending on the context in which the skill was tested, resulting in 15 subscales to be analyzed. In order to accomplish the analysis while maintaining our desired Type I error rate for the eleventh grade testing, we again used a Bonferroni adjustment, splitting the confidence level allocated to Parts 1 and 2 among the fifteen subscales. The Bonferroni adjustment yielded a requirement that each of the fifteen subscales meet a significance level of $.033/15=.0022$ for us to be confident that subscale differences between cohorts were non-chance. We analyzed Part 3 as a single subscale

Draft

with a confidence level of .017.

Each of the fifteen subscales from Part 1 and Part 2, as well as the score on Part 3, was composed of one or relatively few items so that only a limited number of scores were possible. Thus the dependant variables more closely resembled ordinal data, rather than the continuous normally distributed data needed for Ordinary Least Squares Regression and related statistical techniques. For this reason, we analyzed each dependant variable with a Probit analysis.

We used a covariate to control for prior student ability. This covariate was the average of two quantities: 1) each student's sixth grade score on the "quantitative ability" subscale of the Comprehensive Testing Program (CTP) Test published by the Educational Records Bureau; and 2) the student's sixth grade score on reading/writing ability, as measured by the mean of the student's score on the CTP "reading" subscale and the CTP "writing mechanics" subscale.

We did not drop the scores of students who were missing covariate data. Instead, we employed multiple imputation (Rubin, 1987; Mislevy, Beaton, and Sheahan, 1992; Allison, 2002) to impute plausible values for the missing data. The basic idea of multiple imputation is this: given missing data, the ideal approach would be to integrate over all possible values of the missing data, using as your solution the likelihood-weighted average of all the solutions at all these possible values. While performing such an integration is often impossible, a good way to approximate the same process is merely to impute values using a random draw from the probability distribution of what those values are likely to be and to run the statistical analysis with these randomly-selected values. Repeat this process several times and use as your result the average of the several runs. Each randomly selected imputed value is called a "plausible value." Rubin (1987) has demonstrated that using surprisingly few imputations, (i.e., very few sets of plausible values) can provide a very good estimate of the output one would get from an infinite number of imputed values or, equivalently, from actually performing the integration.

Draft

While for most practical applications five plausible values are more than sufficient, we erred on the side of caution, generating fifteen plausible values for each missing data point, using a Markov Chain Monte Carlo process from SAS Proc MI (SAS Institute, 2001). When running the multiple imputation, we supplemented variables used in our final analysis with several additional variables that were highly correlated with students' Grade 6 ability scores, as recommended by Allison (2002). These additional variables were ability scores from Grade 7 (which weren't useful as a covariate because we were unable to meaningfully equate scores from the Traditional cohort to scores from the Block Schedule/IMP cohorts); ability scores from Grade 8 (which were available only for the Traditional cohort and the First Reform cohort); a categorical variable indicating the first high school mathematics class in which the student enrolled; a categorical variable indicating the highest mathematics course the student completed in high school; a set of variables reporting the letter grade the student received in his or her first high school math class; and an indicator of student gender.

We ran each Probit analysis fifteen times (once for each set of plausible values), and we reported as an effect size the average effect size across the fifteen Probit runs. We used SAS Proc Mianalyze (SAS Institute, 2001) to implement formulas reported in Rubin (1987) and Allison (2002) that adjust confidence intervals to account for the greater uncertainty caused by multiple imputation of some missing data. In addition to imputing sixth grade ability scores for some students, on two subscales that contained items both from Part 1 and from Part 2 of the Algebra Achievement Test we used multiple imputation to obtain scores for students who completed one part of the test but were absent when the other part was administered. However, for these two subscales we ran the analysis both using and not using any imputed values, and obtained essentially the same results either way.

We deleted from the analysis the scores of any students who transferred to Suburban
Draft

High School after ninth grade. This left 189 students in the Traditional cohort and 428 students in the Reform cohorts who completed the Part 1 subscales. Also, 192 students in the Traditional cohort and 412 students in the Reform cohorts completed the Part 2 subscales. There were 197 students in the Traditional cohort and 446 students in the Reform cohorts with sufficient data to impute scores for the two subscales combining items from Part 1 and Part 2 (“formulating algebraic models” and “translating the graph of a line into an equation”). Our analysis used scores from 95 pairs of students from the Traditional cohort and 195 pairs of students from the Reform cohorts who completed Part 3 of the Algebra Achievement Test.

Results from Part 1 and Part 2

Results from the first two parts of the Algebra Achievement Test are best described by grouping the items into the nine discrete skills described in Table 1. Table 2 presents results of the probit analysis of these skills. In Table 2, positive effect sizes indicate an advantage for the IMP/Block Schedule cohorts; negative effect sizes reflect the reverse. All p -values and confidence intervals reported in Table 2 are adjusted to account for uncertainty due to multiple imputation.

Table 1. Skills Measured on Algebra Achievement Test 1 and Test 2:

- | |
|--|
| <ol style="list-style-type: none"> 1. formulating an algebraic model (from Test 1); 2. interpreting graphs on the Cartesian plane (from Test 1); 3. translating an exponential function from tabular form to an equation (from Test 1); 4. interpreting an algebraic model (from Test 1); 5. evaluating an expression when it is <ol style="list-style-type: none"> a. presented in context, with access to a graphing calculator (from Test 1) b. presented without context or access to a graphing calculator (from Test 2); 6. solving a linear equation or inequality when it is <ol style="list-style-type: none"> a. presented in context, with access to a graphing calculator (from Test 1) b. presented in standard format, without context or access to a graphing calculator (from Test 2); c. presented in non-standard format involving a ratio, without context or access to a graphing calculator (from Test 2); 7. solving a quadratic equation when it is |
|--|

- a. presented in context, with access to a graphing calculator (from Test 1)
 - b. presented in standard format, without context or access to a graphing calculator (from Test 2);
 - c. presented in non-standard format involving a radical sign, without context or access to a graphing calculator (from Test 2);
 8. simplifying an it expression when it
 - a) have integer coefficients (from Test 2);
 - b) involves ratios (from Test 2).
 9. given a graph of a linear function, writing a linear equation for that function (from Tests 1 and 2)
-

Table 2. Probit Analysis: Effects of Being in the Block Schedule/IMP Cohorts vs. Being in the Traditional Cohort on Skills Tested in Part 1 and Part 2 of the Algebra Achievement Test, after Adjusting for Prior Ability

Skill Tested		Effect Size	Standard Error	p-value	99.78% confidence interval
Formulating an Algebraic Model		+0.32	.09	.0004**	(+.04, +.60)
Interpreting Graphs on the Cartesian Plane		+0.48	.10	<.0001**	(+.17, +.79)
Translating an Exponential Function from Tabular form to an Equation		+0.53	.14	.0002**	(+.10, +.96)
Interpreting an Algebraic Model		+0.08	.09	.3778	(-.20, +.37)
Evaluating Expressions	In Context	+0.08	.11	.4803	(-.25, +.41)
	No Context (note: item used negative #s)	+0.08	.11	.4539	(-.26, +.43)
Solving Linear Equations	In Context	+0.07	.10	.5171	(-.25, +.39)
	No Context	-0.25	.10	.0124	(-.55, +.06)
	No Context with Ratio	-0.08	.12	.5091	(-.45, +.29)
Solving Quadratic	In Context	+0.19	.10	.0560	(-.12, +.50)
	No Context	-0.16	.11	.1241	(-.49, +.16)

Equations	No Context with Radical	0.00	.11	.9895	(-.34, +.34)
Simplifying an Expression	With Integer Coefficients	-0.48	.13	.0002**	(-.88, -.09)
	Involving Ratios	+0.01	.10	.9056	(-.31, +.33)
Translating Graph of a Linear Function into an Equation		-0.48	.10	<.0001**	(-.79, -.17)

** statistically significant ($p < .0022$)

Students who utilized IMP and the semestered block schedule scored significantly higher than did students who used a traditional curriculum and schedule on the three of the nine skills listed in Table 1: formulating a mathematical model to describe a problem situation ($p = .0004$, see Figure 1 and Figure 2); interpreting graphs on a Cartesian plane ($p < .0001$, see Figure 3 and Figure 4); and translating an exponential function from tabular form to an equation ($p = .0002$, see Figure 5 and Figure 6).

Figure 1. Sample Question Requiring Students to Formulate an Algebraic Model

The Watchdog Security Service provides home and business security systems with installation charge of \$150 and a charge of \$5 per week to monitor the system. Write a formula that gives the cost C of the service as a function of time t in weeks that the service is needed.

Figure 2. Mean Score Formulating Algebraic Models

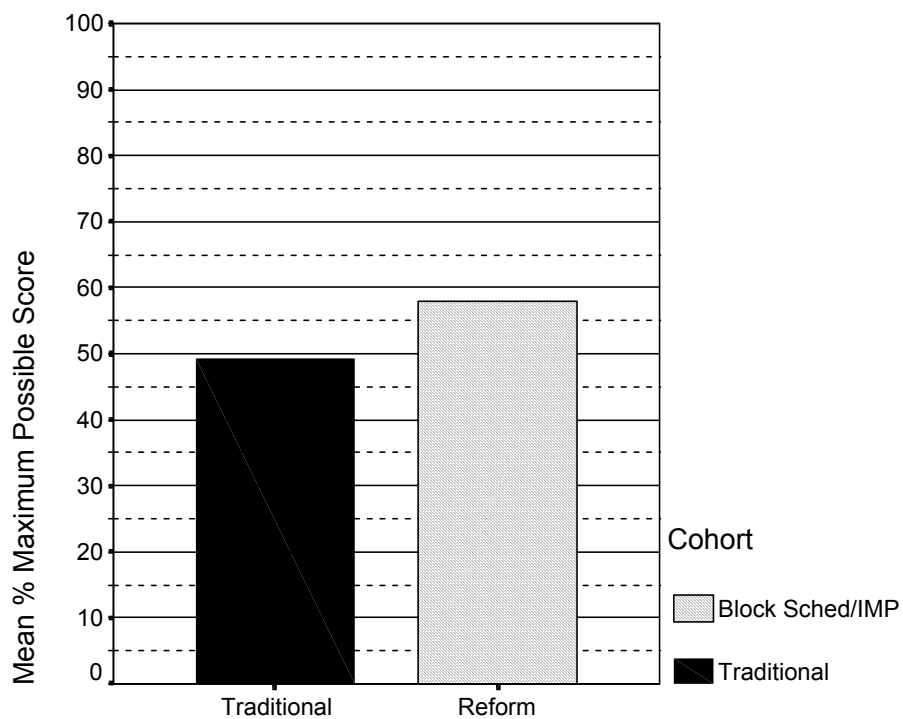
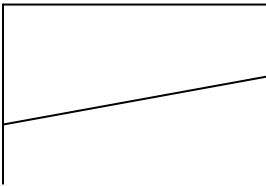


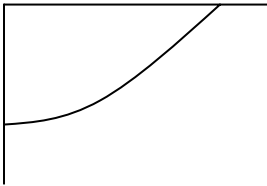
Figure 3. Questions Requiring Students to Interpret Graphs on the Cartesian Plane

Recent reports about bacteria and parasites in city drinking water supplies have led to a booming business in sales of bottled water and water purifying devices. Water has become a big business. The following graphs show three different predictions of how income from water sales might change over the next several years.

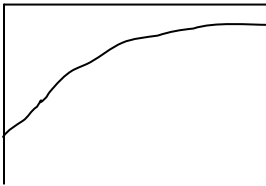
I



II



III



Question 2.1: Which of the three graphs (I, II, or III) shows annual income rising:

- at an increasing rate? _____
- at a constant rate? _____
- at a decreasing rate? _____

Question 2.2: Income from sales of bottled water in 1990 was \$500 million. What is the best match between the three graphs and these equations relating sales income S in millions of dollars to time t in years since 1990?

- $S = 500 + 60t$ _____
- $S = -15t^2 + 290t + 500$ _____
- $S = 500(1.2)^t$ _____

Figure 4. Mean Score Interpreting Graphs on the Cartesian Plane

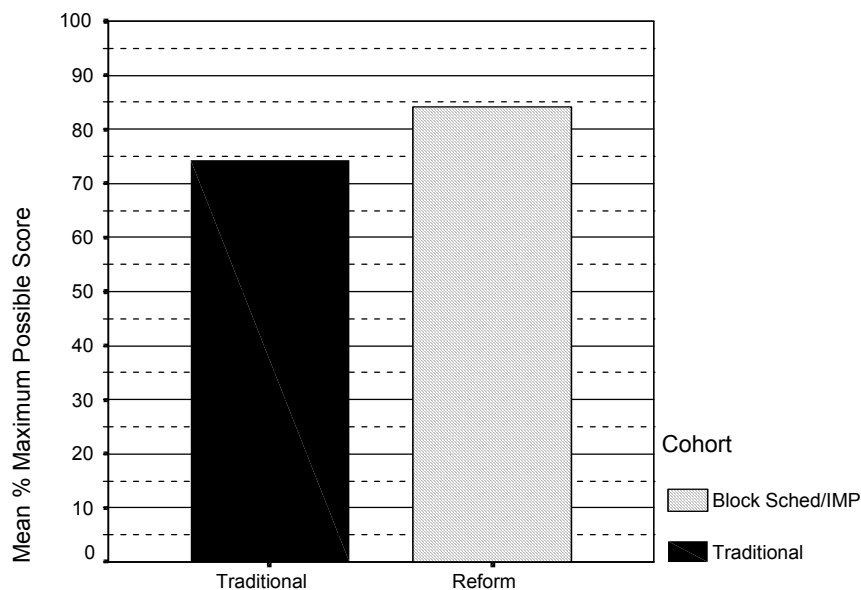
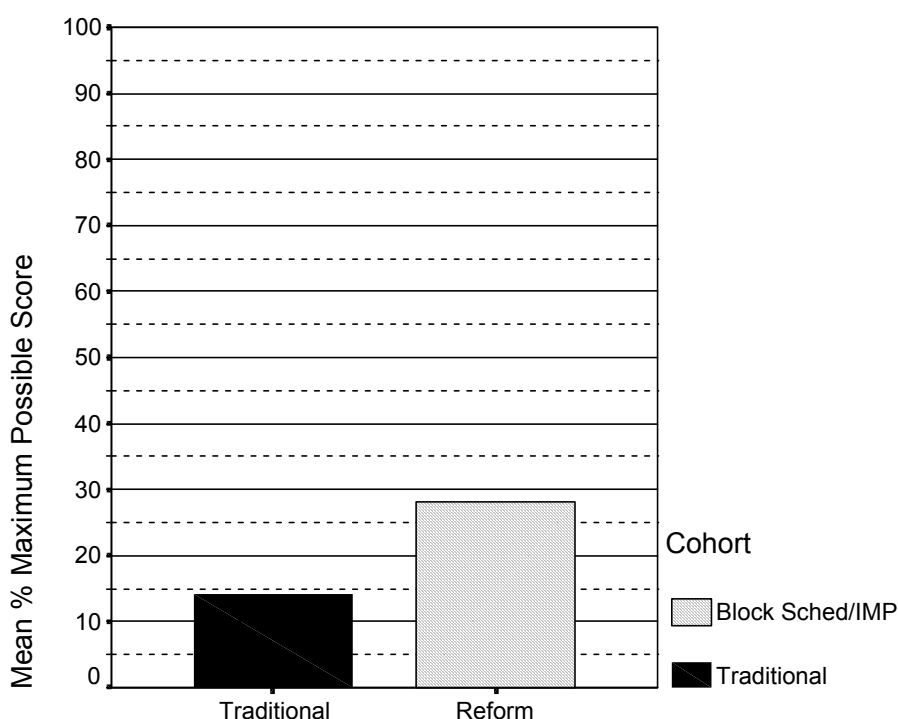


Figure 5. Question Requiring Students to Translate an Exponential Function from Tabular Form to an Equation

Question 1.4: Write an equation relating x and y that will give the pairs of numbers in this table.

x	-1	0	1	2	3
y	0.5	1	2	4	8

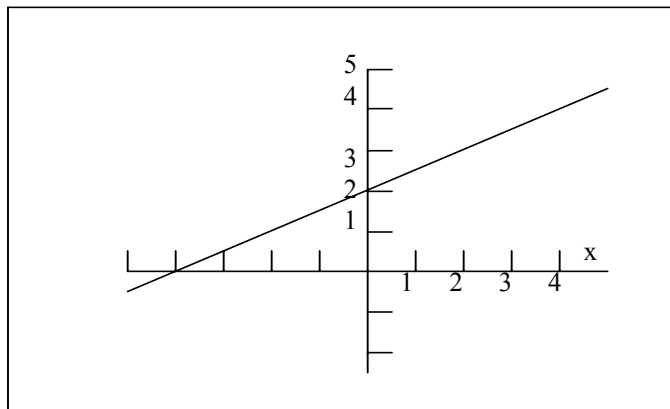
Figure 6. Mean Score Translating an Exponential Function from a Table to an Equation



Students who utilized the traditional curriculum and schedule scored significantly higher than did students who used IMP and the semestered block schedule on one of the nine skills listed in Table 1: translating the graph of a linear function into an equation for that function ($p < .0001$), as shown in Figure 7 and Figure 8. Traditional cohort students also scored significantly higher at simplifying expressions when the expressions had integer coefficients ($p = .0002$) but not when the expressions involved ratios, as shown in Figure 9 and Figure 10.

Figure 7. Questions Requiring Students to Translate the Graph of a Line into an Equation.

Question 1.2: Write an equation relating x and y that will produce this graph.



Question 14: What is an equation of the line shown in the graph below?

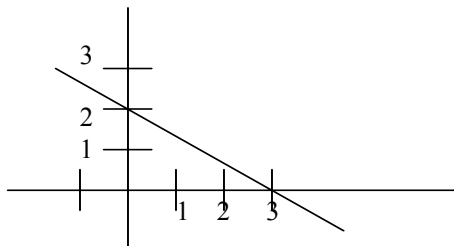


Figure 8. Mean Score Translating a Line Graph to an Equation

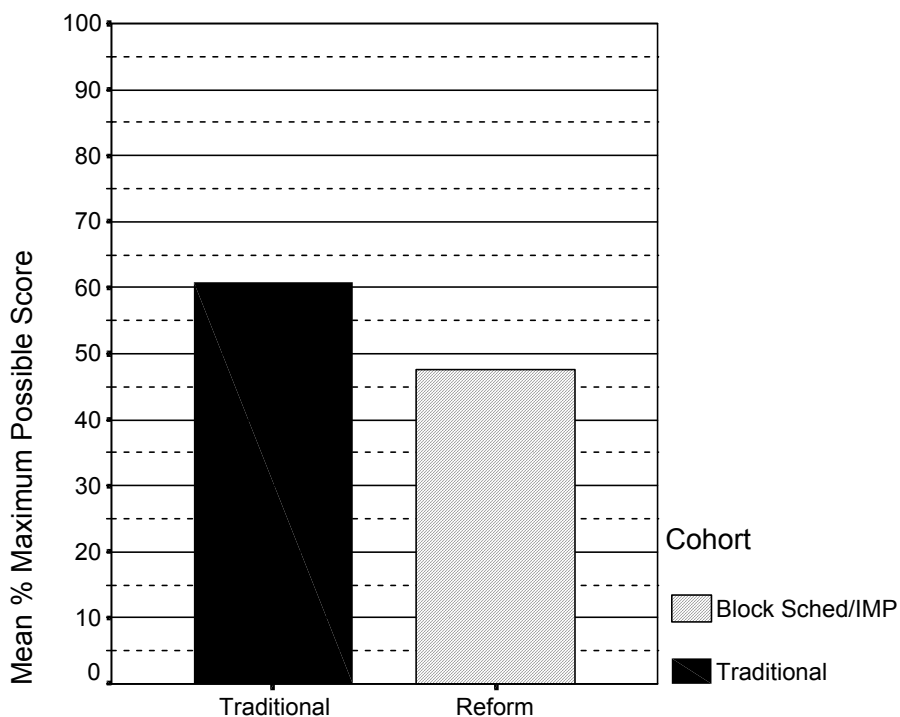


Figure 9. Stems from Multiple-Choice Questions Requiring Students to Simplify Expressions

Expressions with Integer Coefficients

Question 1: Which of the following expressions is equivalent to $(4x^3)(5y)$?

Question 6: Which of the following expressions is equivalent to $-3(5x - 6)$?

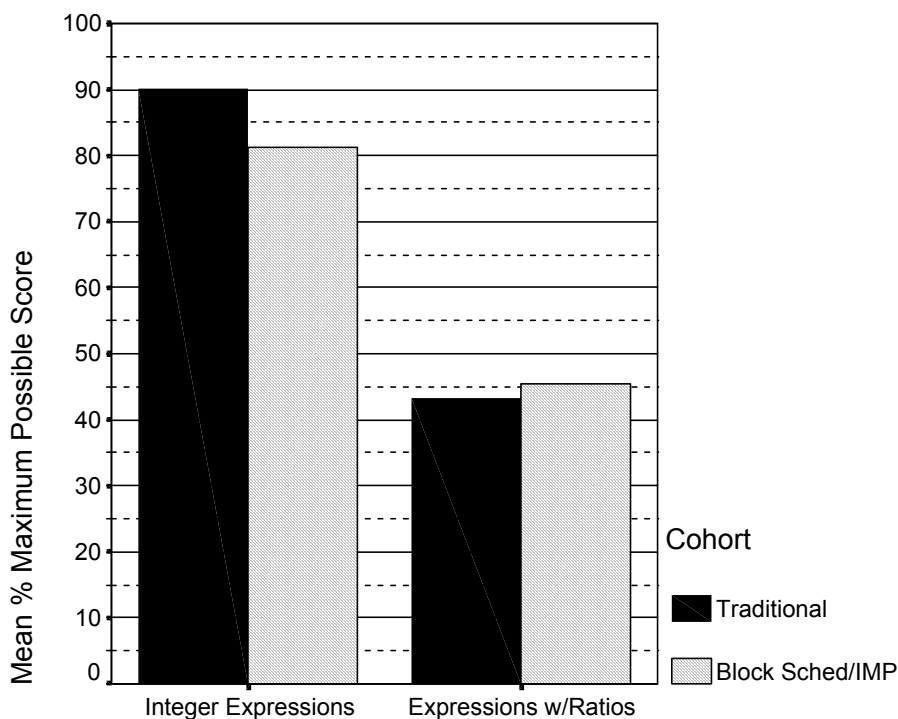
Expressions Involving Ratios

Question 2: Which of the following expressions is equivalent to $\frac{125+x}{25}$?

Question 4: If $t \neq 0$, which of the following expressions is equivalent to

$$\frac{8}{t} - \frac{5}{2t} \quad ?$$

Figure 10. Mean Score Simplifying Expressions



Differences between students in the Reform cohorts and students in the Traditional cohort on the remaining four subskills listed in Table 1 and Table 2 are displayed in Figures 11-14. Differences between groups on these skills were not statistically significant. However, all four figures show a trend for students in the Reform cohorts to score higher than students in the Traditional cohort when items were presented in context. When symbol manipulation items were presented without context and in a standard format, students in the Traditional cohort sometimes scored higher than students in the Reform cohorts, as can be seen in Figure 13 and Figure 14. However, if context-free symbol manipulation items involved multiplication by negative numbers (Figure 12), or were in a non-standard format involving ratios (Figure 13), or involving radicals (Figure 14), the two groups scored about equally.

Figure 11. Mean Score Interpreting Algebraic Models

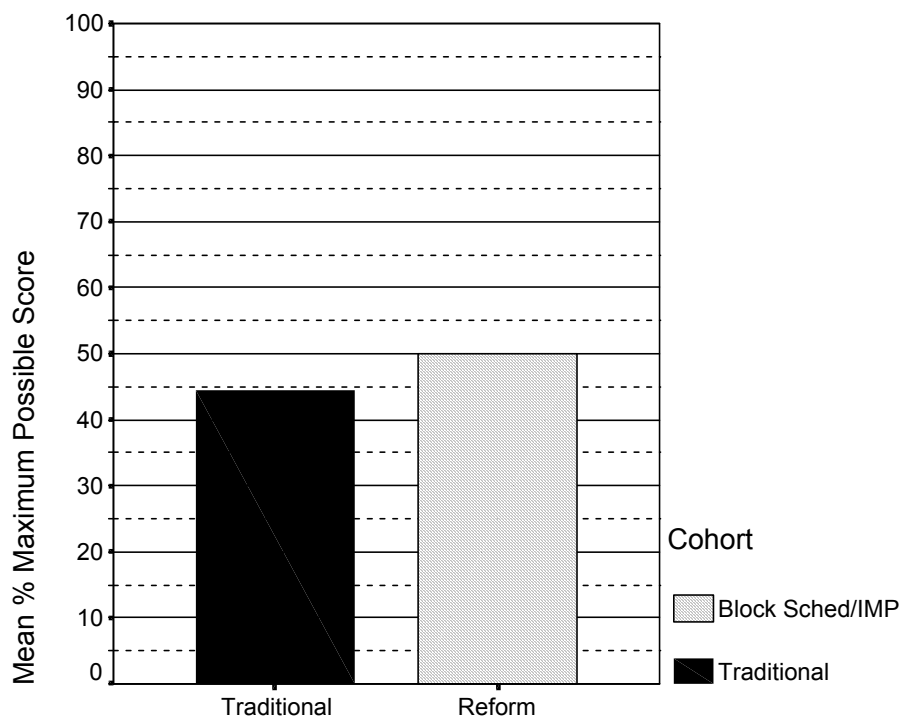


Figure 12. Mean Score Evaluating Expressions

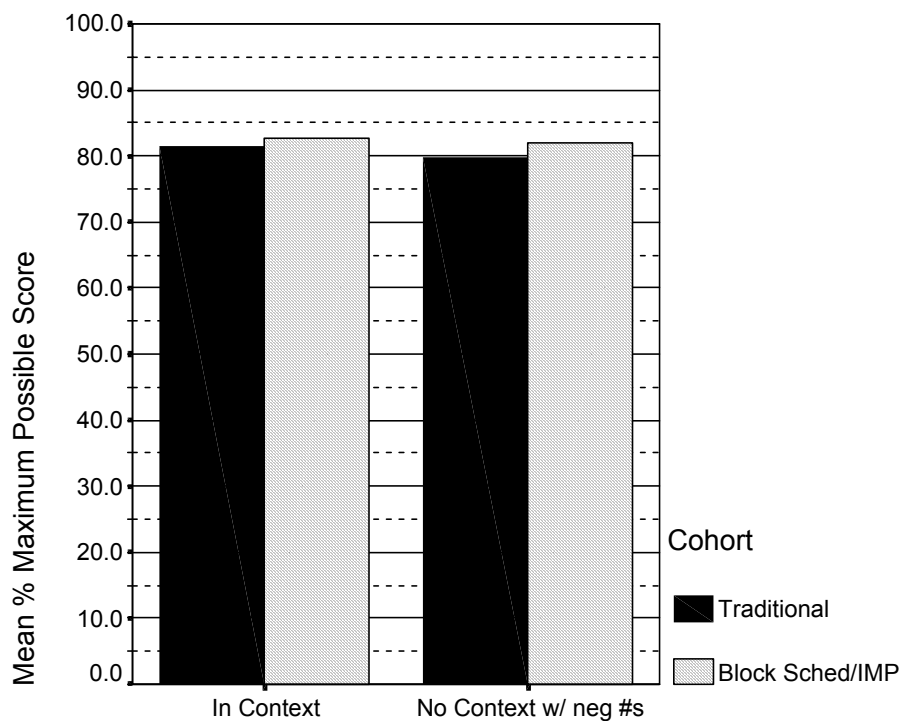


Figure 13. Mean Score Solving Linear Equations

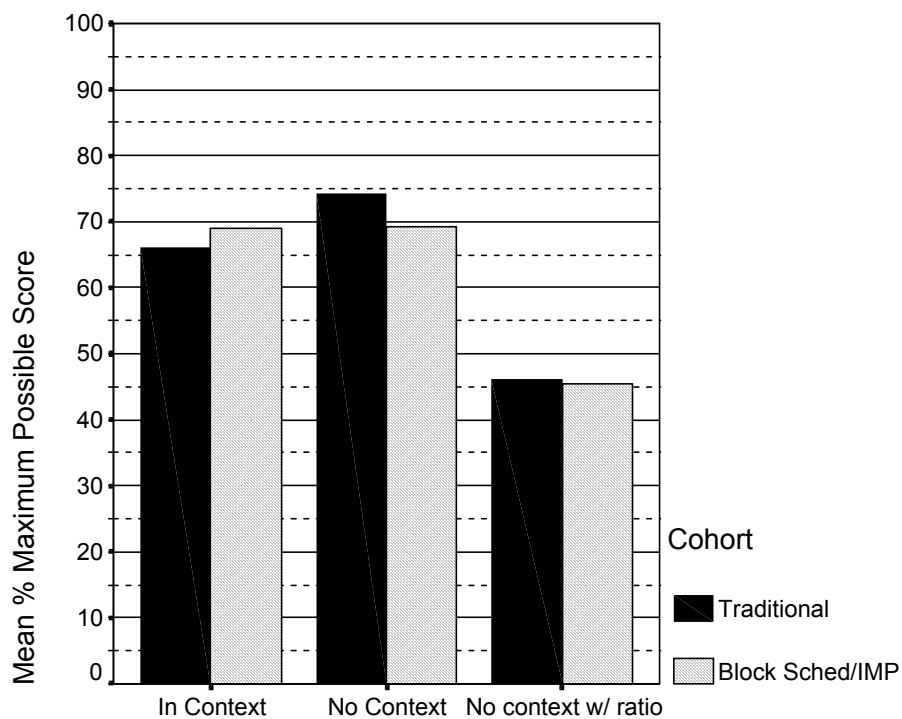
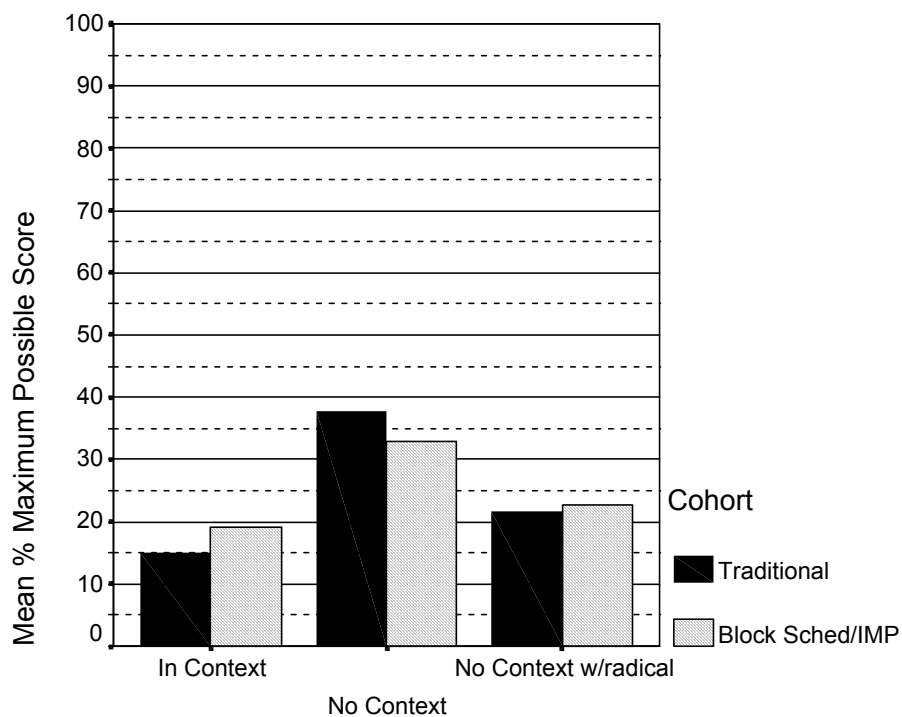


Figure 14. Mean Score Solving Quadratic Equations



Results from Algebra Achievement Test Part 3.

Part 3 of the Algebra Achievement Test required students working in pairs to apply multiple algebraic skills to solve the extended, applied algebra problem shown in Figure 15.

Figure 15. Algebra Achievement Test Part 3: Extended, Applied Algebra Problem.

Many companies sell their products through long-distance telephone calls to customers. For example, the *CD Club* sells music compact discs all across the country from its headquarters in New York. Sales calls made by the *CD Club* last an average of about 4 minutes apiece. The club has bids from three possible providers of their long-distance telephone service:

- 1) *Apple Communications* will charge \$0.35 for placing each call and then \$0.15 per minute of time used in the call.
- 2) *Bell Telephone* makes a fixed charge of \$260 per week for access to its long-distance lines, but charges only \$0.10 per minute of time used by the calls.
- 3) *Capital Long Distance Services* will make a fixed charge of \$600 per week for unlimited use of its long-distance lines.

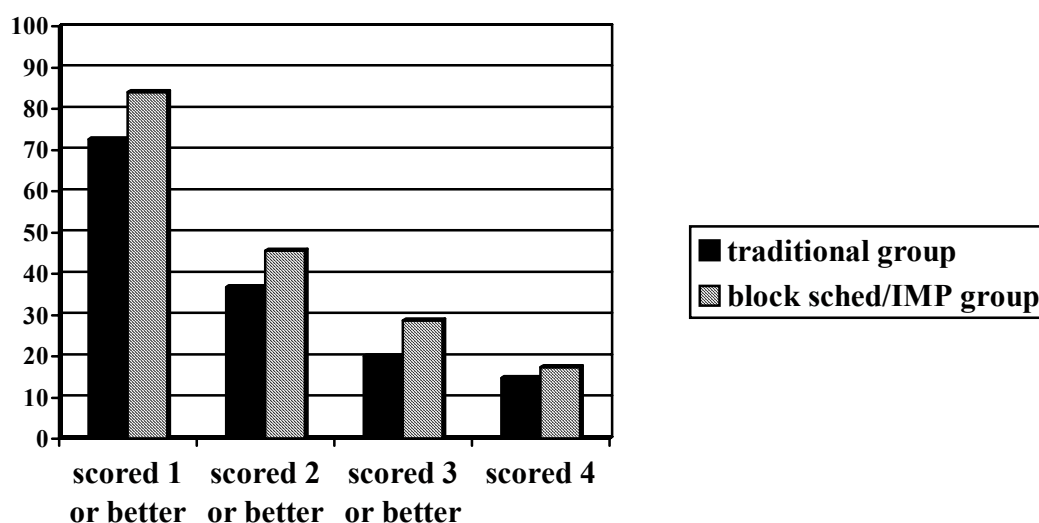
Question: For the *CD Club* the problem is to choose the long-distance service that is least expensive for their business. What advice would you give about the phone company to choose and why?

To solve the problem correctly, pairs of students had to perform the following subtasks:

1. realize that the “best” choice of provider would depend on how many calls the business made per week;
2. determine a correct algebraic equation for the weekly cost of using each provider, as a function of number of calls per week;
3. correctly solve the equations;
4. provide a clear and complete explanation to the business about which company it should choose and under what conditions.

Pairs of students were assigned scores of 0 through 4, depending on how many of these subtasks that had performed correctly. Figure 16 displays the percent of students in the IMP/Block Schedule group and in the Traditional group who scored at each level on Test 3.

Figure 16. Percent Scoring at Each Level on Test 3, an Applied Algebra Task.



As can be seen in the figure, pairs in the Block Schedule/IMP cohorts tended to score higher than those in the Traditional cohort. A Probit analysis (with multiply imputed values for missing ability scores) estimated that after controlling for prior ability, membership in the IMP/Block Schedule group accounted for an improvement of .32 standard deviations in a pair's score, with a 98.3% confidence interval for the effect size of $(-.02, +.66)$, $p=.023$, which approaches statistical significance. (As noted previously, a Bonferroni adjustment to keep the overall confidence level at .05 for Parts 1, 2, and 3 of the eleventh grade Algebra Achievement Test would require a significance level of .017 for Part 3. The .023 confidence level equates to an overall p -value of $3*.023 = .069$ for the 3-part Algebra Achievement Test.)

Analysis of Grade 12 NAEP Blocks

Methodology

We obtained the three NAEP blocks completed by twelfth graders at Suburban High School from the National Center for Educational Statistics website in the spring of 1998. Unfortunately, the actual NAEP forms we utilized had no specific place where students were asked to put their names. A number of teachers administering the test in 1998 misconstrued this to mean that students should not put their names on their test forms. This created two difficulties in analyzing the data.

First, because each of the three blocks was completed and stored separately, there was no way to link the three blocks together to create overall scale scores across the three forms. For this reason, we compared the Traditional and IMP/Block groups item by item, rather than on overall subscales. This resulted in 28 statistical comparisons. (There were 29 items across the three tests. We dropped item 4 from block 1996-M10 from the analysis because the NAEP scoring guide could only give valid results under the assumption that students completed the item without having a ruler available, and in all years students at Suburban High School were administered the test with rulers available.) For each of the 28 items, we used a Probit analysis to compare the achievement of students in the Traditional cohort to that of students in the Block Schedule/IMP cohorts.

There were some students who stopped working on one or more NAEP blocks before completing that section of the test. In such cases, items subsequent to the first blank item within a block were coded as “not reached” instead of “incorrect.” Coding such items as “not reached” ensured that only scores from students who actually had a chance to address each item were utilized in comparing the achievement of students in the Traditional cohort to that of students in the Block Schedule/IMP cohorts. As a separate indicator of student mathematics knowledge, we

used a Chi-square cross tabulation for each NAEP block to analyze whether there was any difference between groups in the likelihood of a student failing to finish the test.

Combined with the 28 comparisons for individual items, analyzing “likelihood of leaving the test unfinished” for each block yielded a total of 31 statistical comparisons. Keeping the Type I error rate on the analysis of NAEP items at 5% necessitated using a very strict confidence level of $.05/31=.0016$ for us to be confident that the difference between cohorts on any given item was non-chance.

A second problem caused by the lack of names on many of the tests completed in 1998 was that student scores could not be linked to student ability, as measured by scores on mathematics achievement tests completed prior to high school. Because only the subset of students in each cohort who were enrolled in math class during the second semester of their senior year completed the NAEP blocks, it was important to demonstrate that the Traditional cohort students tested were in fact comparable to the Reform cohort students tested. We accomplished this by comparing the mean Grade 6 score on CTP Quantitative Ability for the pool of seniors enrolled in mathematics class in the spring of 1998 to that of the pool of seniors enrolled in mathematics class in the spring of 2001 and the spring of 2002. We expressed these scores in terms of standard deviations above the national mean. Grade 6 scores were available for 93 of the 138 seniors enrolled in spring 1998, with a mean of 1.18. Grade 6 scores were available for 97 of the 133 seniors enrolled in the spring of 2001, with a mean of 0.86 and for 122 of the 169 seniors enrolled in spring, 2002 with a mean 1.18. These scores all reflect the very high-achieving population attending Suburban High School. Students in the Traditional group completing the NAEP test may have had an advantage in initial ability over students in the First Reform cohort, and appear to have been comparable to those in the Second Reform cohort who took the test.

Results

Table 3 describes the results of the Probit analysis on the 28 NAEP items completed by students at the end of twelfth grade. Positive effect sizes indicate an advantage for the IMP/Block Schedule cohorts; negative effect sizes reflect the reverse.

NAEP categorizes items from the three blocks used for this study in two ways. Each item is assigned to one of five content classifications: “Number sense, properties, and operations,” “Measurement,” “Geometry and spatial sense,” “Data analysis, statistics, and probability,” or “Algebra and functions.” Each item is also assigned to one of three mathematical abilities tested: “Problem solving,” “Conceptual understanding,” or “Procedural knowledge.” Table 3 reports results according to these classifications. For purposes of this study, we combined the classification of “Geometry and spatial sense” was combined with “Measurement” and we combined “Algebra and functions” with “Number sense.”

Table 3 refers to individual NAEP items using a shorthand based on designations for the NAEP blocks that were used in earlier versions of the NAEP website. Block 1996-12M10 is referred to as “Block G,” block 1996-12M12 is referred to as “Block C” and Block 1996-12M13 is referred to as “Block P.” For example, the item labeled “c2: apply average” is the second item in block C (i.e., Block 1996-12M12), which requires students to apply the concept of average.

Table 3 Probit Analysis: Effect of being in a Block Schedule/IMP Cohort vs. Being in the Traditional Cohort on Student Scores for 28 NAEP Items

Skill Tested			Effect Size	Stand. Error	p-value	99.84% confidence interval
	Problem Solving	c2: apply average	+0.29	.14	.0341	(-.14,+.72)
		c10: compare mean and median	+0.46	.14	.0007**	(+.03,+.89)

Probability and Statistics		g5: use data from a chart	+0.32	.18	.0708	(-.24,+.87)
		p5: identify superfluous information	+0.31	.17	.0630	(-.22,+.84)
	Concepts	c9: compare probabilities	+0.49	.15	.0010**	(+.02,+.96)
		p6: combinatorics	+0.75	.18	<.0001**	(+.19,+1.32)
	Procedures	p1: read graph	+0.48	.20	.0179	(-.16,+1.11)
Geometry and Measure- ment	Problem Solving	g2: create shape	+0.61	.16	.0002**	(+.10,+1.12)
		g3: prove equiv. areas	+0.58	.16	.0002**	(+.09,+1.07)
		g10: find disc center	+0.38	.14	.0077	(-.07,+.82)
		p4: find circumference	+0.55	.16	.0004**	(+.06,+1.03)
		p7: measure degrees (circle)	+0.22	.16	.1580	(-.27,+.71)
		p8: measure degrees (angle)	+0.01	.16	.9244	(-.48,+.51)
	Concepts	c3: choose unit of measure	+0.42	.22	.0534	(-.27,+1.11)
		c6: compute cylinder volume	+0.29	.14	.0340	(-.14,+.73)
		g1: compare shapes	+0.34	.23	.1417	(-.40,+1.08)
	Procedures	c1: use similar triangles	+0.06	.15	.6766	(-.42,+.55)
		p2: draw parallelogram w/perpendicular diagonals	-0.36	.16	.0221	(-.85,+.14)
	Problem Solving	c5: use percent increase	+0.34	.14	.0131	(-.09,+.77)
		p9: describe algebraic pattern	+0.25	.14	.0688	(-.18,+.67)

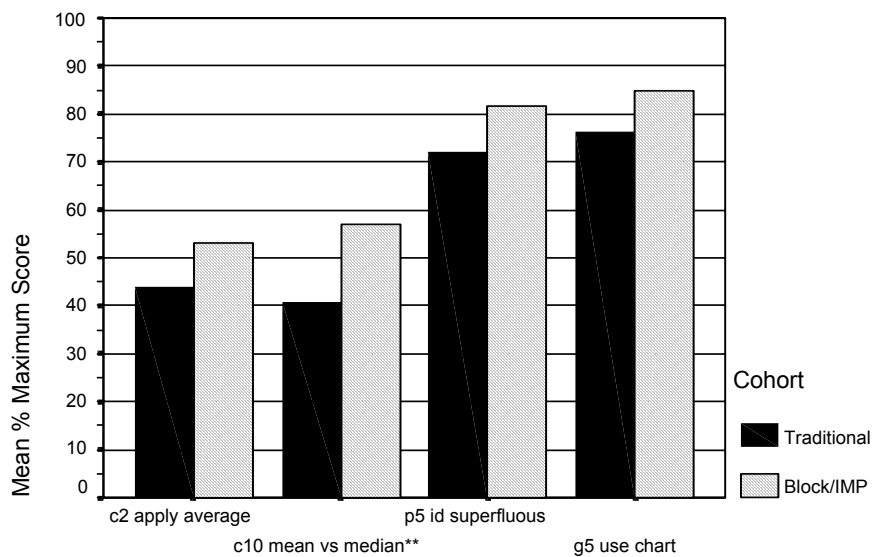
Algebra and Number Concepts	Concepts	c4: numerical reasoning (odd/even)	+0.04	.19	.8231	(-.56, +.65)
		c7: divide large numbers	+0.09	.16	.5446	(-.41, +.61)
		g6: use zero principle	+0.66	.24	.0052	(-.09, +1.40)
		g7: algebraic reasoning (odd/even)	+0.18	.16	.2649	(-.33, +.69)
		g9: identify graph of function	+0.20	.16	.2272	(-.32, +.71)
	Procedures	c8: use ratios in context	+0.46	.14	.0014**	(+.01, +.92)
		p3: solve rate vs. time problem	+0.38	.16	.0164	(-.12, +.89)
		g8: use trigonometric identity	+0.37	.16	.0217	(-.14, +.87)

** statistically significant ($p < .0016$)

As Table 3 shows, students who utilized the semestered block schedule and IMP core curriculum scored higher on 27 of the 28 NAEP items tested than did students who utilized a traditional schedule and traditional core curriculum. On seven of the items tested, the advantage of the Block Schedule/IMP students met our very stringent .0016 criterion for statistical significance.

Figures 17-25 report the difference between the Traditional cohort and the Block Schedule/IMP cohorts for each of these 27 items in terms of the average percent of the maximum possible score achieved by each group. (Note that the majority of items were scored on a 0/1 scale. For 0/1 items “mean percent of maximum possible score” is simply the percent of students in the cohort who got the item correct.)

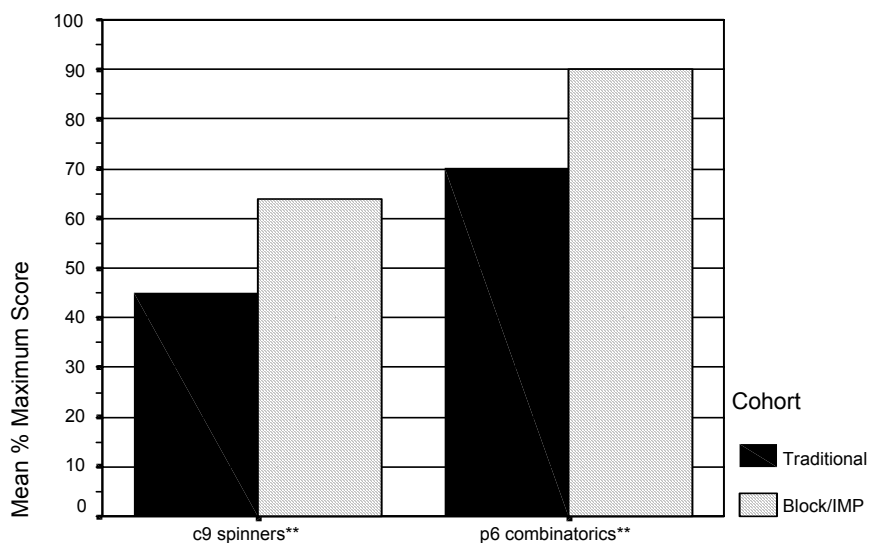
Figure 17. Problem Solving in Probability, Statistics, and Data Analysis Content Area



Probab/Stats/Data PROBLEM SOLVING items

** item is statistically significant ($p < .0016$)

Figure 18. Conceptual Understanding in Probability, Statistics, and Data Analysis Content Area



Prob/Stats/Data CONCEPTUAL UNDERSTANDING items

** item is statistically significant ($p < .0016$)

Figure 19. Procedural Item in Probability, Statistics, and Data Analysis Content Area

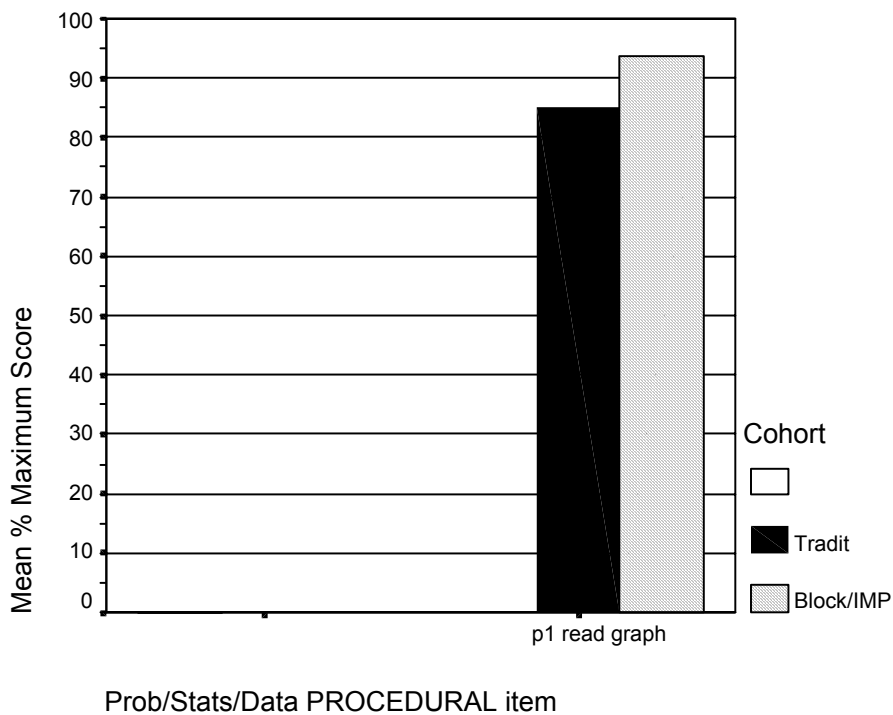
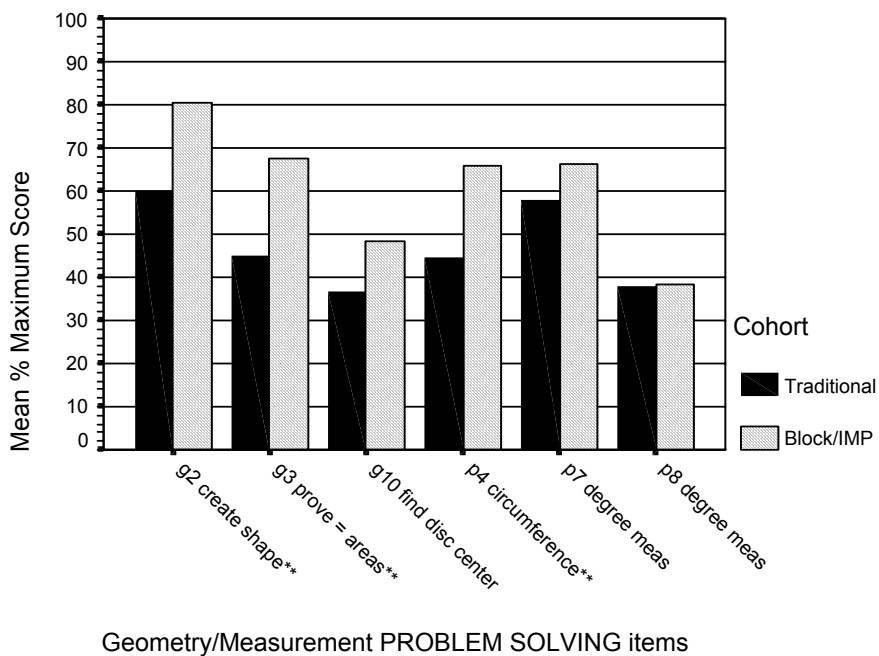


Figure 20. Problem Solving in Geometry and Measurement Content Areas



** item is statistically significant ($p < .0016$)

Figure 21. Conceptual Understanding in Geometry and Measurement Content Areas

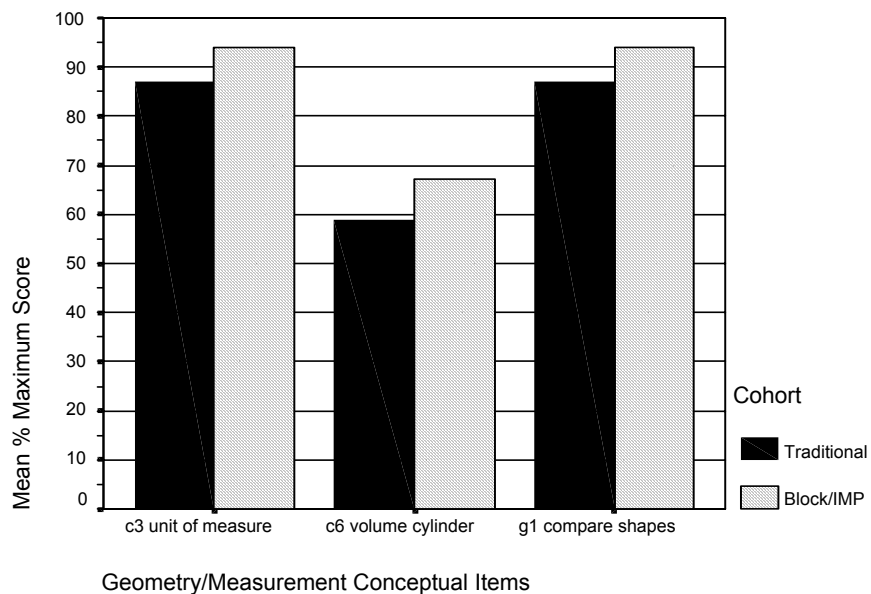


Figure 22. Procedural Items in Geometry and Measurement Content Areas

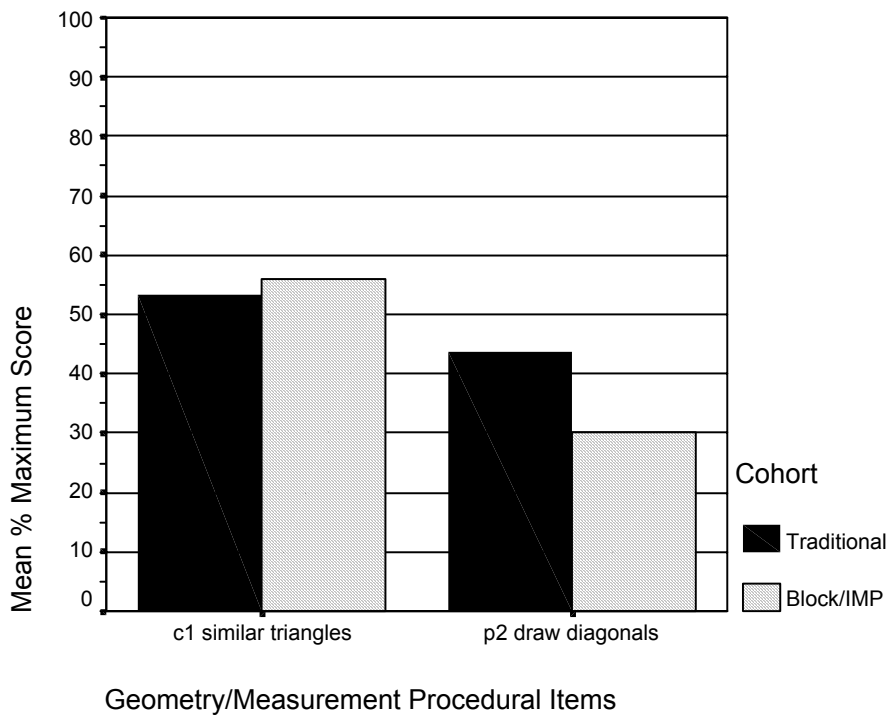


Figure 23. Problem Solving Items in Algebra and Functions or Number Content Areas

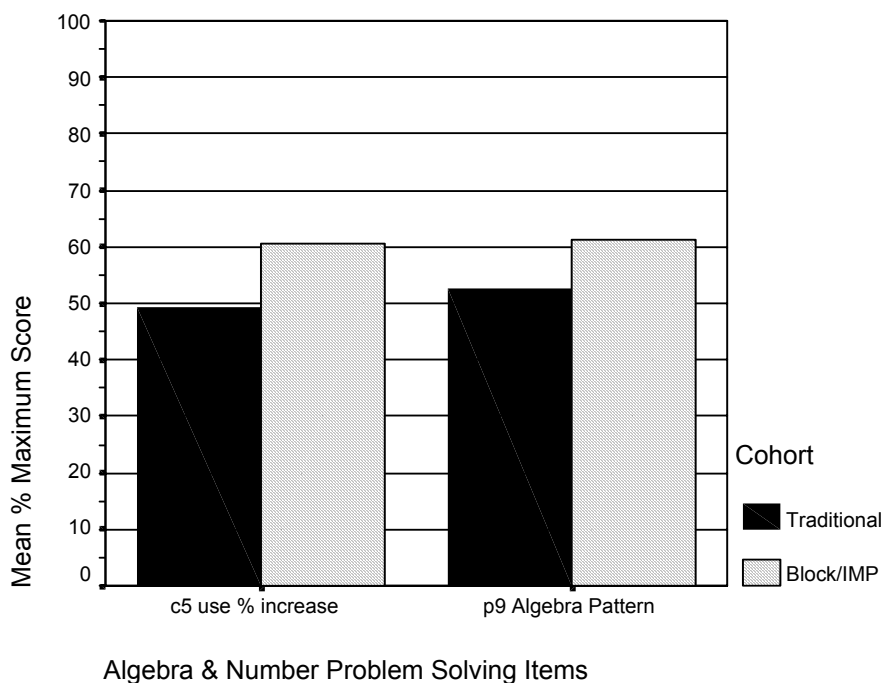


Figure 24. Conceptual Items in Algebra and Functions or Number Content Areas

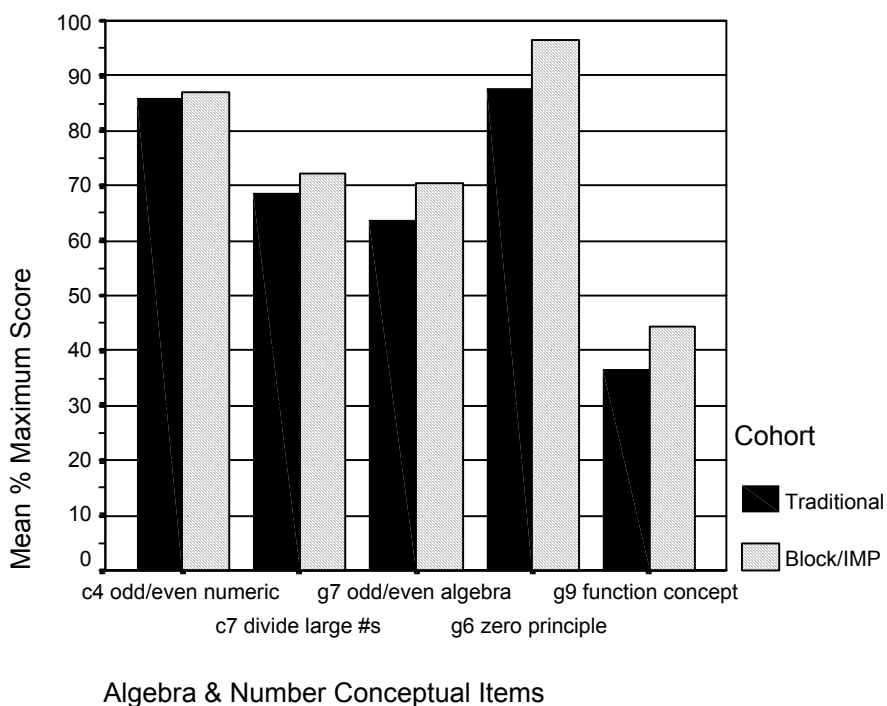
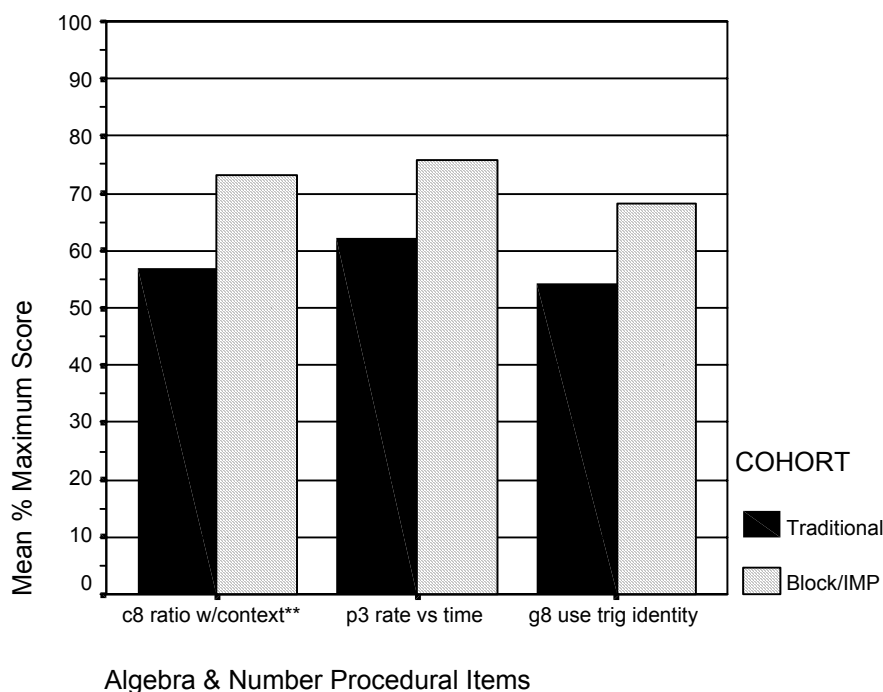


Figure 25. Procedural Items in Algebra and Functions or Number Content Areas



** Item statistically significant ($p < .0016$)

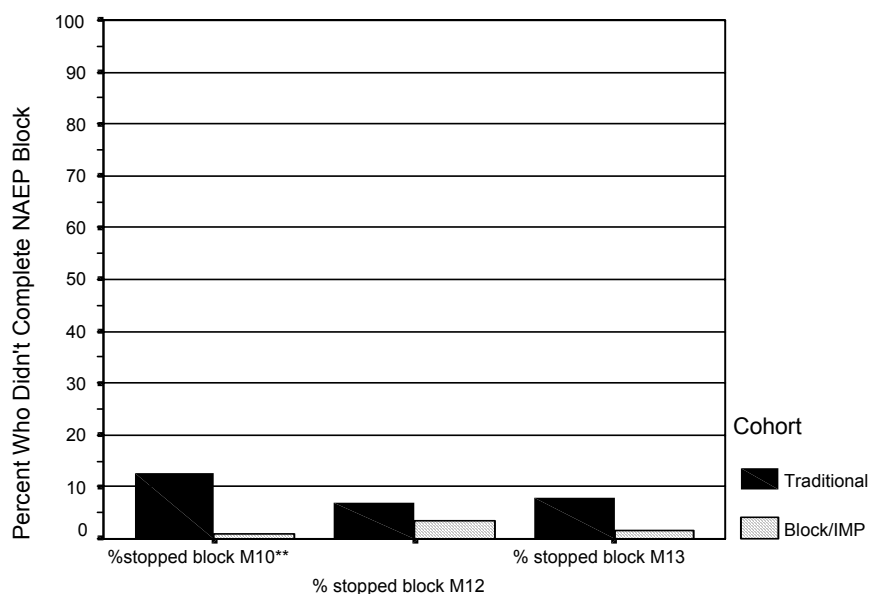
Likelihood of giving up before completing NAEP blocks. We did not limit the amount of time students had for completing the three NAEP blocks. Nonetheless, some twelfth graders stopped answering questions on one or more blocks of the test before completing it. While these students were not marked wrong on items they did not reach, “giving up” is itself likely to be an indication either of difficulty answering the problems or of dislike for mathematics tasks.²

As shown in Figure 26, 12.7% (13 of 102 students) in the Traditional cohort failed to

² Had we marked not-reached items “incorrect,” the advantage of the Block Schedule/IMP cohorts would have reached statistical significance on two additional items: item g6, which required students to apply the zero principle to solve simultaneous algebraic equations (modified effect size = 0.75 standard deviations; $p = .0001$); and item g10, which required students to use a geometric construction to find the center of a disc (modified effect size = 0.51 standard deviations; $p = .0010$).

finish NAEP Block G (1996-12M10), vs. only 1% (2 of 200 students) in the Block Schedule/IMP cohorts. This difference was statistically significant ($p < .0001$). Traditional cohort students were also more likely to leave the other NAEP blocks unfinished, although the contrast did not reach our .0016 criterion for statistical significance. Thus, 6.9% (7 of 101 students) in the Traditional cohort failed to finish NAEP Block C (1996-12M12), vs. only 3.5% (7 of 200 students) in the Block Schedule/IMP cohorts ($p = .182$). Finally, 7.8% (8 of 102 students) in the Traditional cohort failed to finish NAEP Block P (1996-12M13), vs. only 3.5% (3 of 203 students) in the Block Schedule/IMP cohorts ($p = .0049$).

Figure 26. Percent of Students Failing to Complete NAEP blocks by Cohort



** Difference between cohorts statistically significant ($p < .0016$)

Comparison of Honors students in the Pilot and Reform Cohorts

Methodology

In addition to testing seniors in the Traditional cohort during the spring of 1998 and seniors in the Block Schedule/IMP cohorts during the springs of 2001 and 2002, we administered Draft

the three NAEP blocks to seniors in a Pilot cohort during the spring of 2000. Like students in the Block Schedule/IMP cohorts, students in the Pilot cohort utilized a semestered block schedule throughout high school. Unlike the Block Schedule/IMP cohorts, in the Pilot cohort Honors students used the traditional Algebra 1, Geometry, Algebra 2 sequence, while other students utilized four IMP courses as their core sequence. The NAEP blocks completed by Honors students in the Pilot cohort enabled us to address an important question: For this top-ability group, did achievement under a block schedule change when the school switched from a traditional core mathematics curriculum to IMP?

Fortunately, student tests completed in 2000 and thereafter contained each student's name, so it was possible to link student scores to prior ability, as measured by scores in mathematics achievement tests completed prior to high school. As a covariate for this analysis, we used the mean of each student's seventh grade score on the "Mathematics" and "Quantitative Ability" subscales of the CTP test administered by the Educational Records Bureau. These seventh grade scores, which for reasons described in Kramer (2002) had been unavailable to use in analyzing results of the Algebra Achievement Test, provided a better covariate than did the sixth grade scores we used in other analyses.

In the Pilot cohort, 34 Honors students completed the three NAEP blocks. In the Block Schedule/IMP cohorts, 63 Honors students completed all three NAEP blocks. In addition, 5 Honors students in the Block Schedule/IMP cohorts completed at least one NAEP block but were absent for part of the testing due to a school-sponsored activity.

Due to the relatively small sample size, we deemed it unlikely that we would be able to detect meaningful differences if we utilized a Bonferroni adjustment with a large number of subscales. Therefore, we decided to compute only three aggregate scales, one for each of the "Mathematical Abilities" identified by the NAEP authors: "Problem Solving," "Conceptual
Draft

Understanding,” and “Procedural Knowledge.” In testing each scale, we used a Bonferroni-adjusted confidence level of .017, thus maintaining an overall $.017 \times 3 = .051$ confidence level for our investigation of Honors students’ achievement.

Honors students tended to cluster at a relatively small number of scores on the three subscales we analyzed, with a large ceiling effect for both the “Conceptual understanding” and for the “Procedural knowledge” scales. Therefore, we treated the data as ordinal, employing a Probit analysis. For the “Problem solving” subscale, which contained responses more closely resembling normal data than did the other two, we checked our results by performing an Ordinary Least Squares Regression, which obtained substantially the same results as the Probit analysis.

As we had in analyzing the Algebra Achievement Test, we utilized Multiple Imputation to create fifteen plausible values for students with missing ability scores. We also used the partial information available to impute plausible subscale scores for the five students who completed only part of the test. In addition to seventh grade ability and NAEP block scores, we utilized the following variables for the multiple imputation: Grade 6 CTP scores, Grade 8 CTP scores (which were available only for students in the Pilot cohort and First Reform cohort); a categorical variable indicating the first high school mathematics class in which the student enrolled; a set of variables reporting the letter grade the student received in his or her first high school math class; and an indicator of student gender. As with the eleventh grade data, we utilized SAS Proc MI (SAS Institute, 2001) to perform the multiple imputation, and we utilized SAS Proc Mianalyze (SAS Institute, 2001) to adjust confidence intervals to account for the greater uncertainty caused by multiple imputation of some missing data.

Results

Changes in number of courses completed. It is important to note that in replacing a core curriculum consisting of three traditional courses (Algebra 1/Geometry/Algebra 2) with a core IMP curriculum, Suburban High School implemented two innovations that might improve achievement under a semestered block schedule: 1) they adopted a curriculum that emphasized in-depth student investigations while de-emphasizing lecture, and 2) they spread mathematics topics over a larger number of courses, replacing three core courses with four core courses. The latter adaptation resulted in Honors students completing more courses during high school. On average, Honors students in the Block Schedule/Traditional Curriculum group who took the NAEP test completed 5.1 semesters of mathematics over four years of high school. In contrast, the average Honors student in the Block Schedule/IMP Curriculum group who took the NAEP test completed 5.6 semesters of mathematics over four years of high school. The difference in the number of courses completed may be partly responsible for the achievement results reported in the next section.

Changes in achievement. Table 4 compares Honors students who used a traditional core curriculum to Honors students who used the IMP curriculum on each “Mathematical Ability” measured by the NAEP blocks. Both groups used a semestered block schedule throughout high school. Positive effect sizes indicate an advantage for the IMP curriculum group; negative effect sizes reflect an advantage for the Traditional Curriculum group. As Table 4 shows, Honors students using the IMP curriculum had a substantively and statistically significant advantage of 0.57 standard deviations on NAEP conceptual items. The sample size of Honors students was not sufficient to be certain that differences on Problem Solving or Procedural items were non-chance.

Table 4 Probit Analysis: Effect on Honors Student Achievement Under a Semestered Block

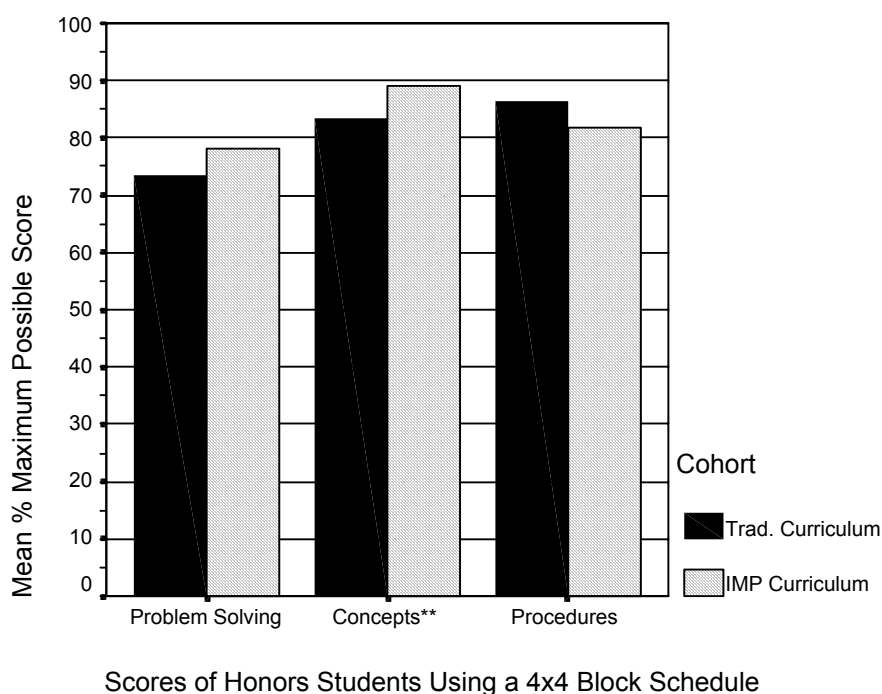
Schedule of Using the IMP Curriculum vs. a Traditional Core, after Adjusting for Prior Ability

Mathematical Ability Tested	Effect Size	Standard Error	<i>p</i> -value	98.3% confidence interval
Problem Solving	+0.31	.22	.155	(-.21, +.83)
Conceptual Understanding	+0.57	.23	.014**	(+.02, +1.12)
Procedures	-0.38	.23	.104	(-.94, +.18)

** exceeds Bonferroni-adjusted standard of $p=.017$ for non-chance results

Effect sizes in Table 4 are adjusted for prior ability. Figure 27 supplements this information by contrasting the raw scores of Traditional with IMP Honors students on each “mathematical ability” scale.

Figure 27. Scores on NAEP “Mathematical Abilities” Subscales, Grade 12 Honors Students



** statistically significant ($p<.0017$)

Changes in Mathematics Course Enrollment: Traditional Cohorts vs. Reform Cohorts

Methodology

We had access to an automated database containing complete high school transcripts of students at Suburban High School going back to the class who graduated in the spring of 1996. In analyzing the effects of the new schedule and curriculum on student course enrollment, we decided not to use the graduating class of 1998 as a “Traditional” comparison group. Though they were the only Traditional group who had completed the Algebra Achievement Test and the NAEP blocks, they had utilized a semestered block schedule during their senior year. For this reason, we felt they would not provide an accurate picture of course-taking patterns at Suburban High School under a traditional schedule and curriculum. Therefore, we decided to use students from the two graduating classes prior to the graduating class of 1998 as “Traditional” comparison cohorts.

We analyzed the transcripts of all students who were enrolled at Suburban High School throughout Grades 9-12, and who were in the graduating class of 1996, the graduating class of 1997 (the Traditional cohorts), the graduating class of 2001, or the graduating class of 2002 (the Reform cohorts). Students who transferred to Suburban High School subsequent to the first semester of their freshman year, or who did not stay through the fourth semester of their senior year, were dropped from the database. This left a sample of 209 students who graduated in the spring of 1996, 207 students who graduated in the spring of 1997, 234 students who graduated in the spring of 2001, and 263 students who graduated in the spring of 2002.

We counted students as having been enrolled in a mathematics class only if their transcripts indicated they had passed the class and received credit. Under the traditional schedule, each course was allocated 129 hours of class time. In 1998-99, when students in the graduating class of 2001 (the First Reform cohort) were in ninth grade, each course was allocated

Draft

106.67 hours. Subsequent to 1998/99, each course was allocated 120 hours. There were fewer allocated hours in 1998/99 because that year Suburban High School was piloting a form of semestered block schedule that allocated 80 days per semester and reserved a 20-day special spring session for in-depth projects. (Usually, the in-depth projects were in subject areas other than mathematics.) Subsequent to 1998/99 the twenty-day spring session was discontinued.

To compute the number of hours students were enrolled in mathematics classes we multiplied the credits earned in each class times the number of allocated hours for that class, and summed this value over all mathematics classes the student took. All mathematics classes were worth one credit, except for a course taken by some students in the Reform cohorts entitled “SAT Prep Math,” which was taken simultaneously with “SAT Prep Verbal” and was worth one half mathematics credit.

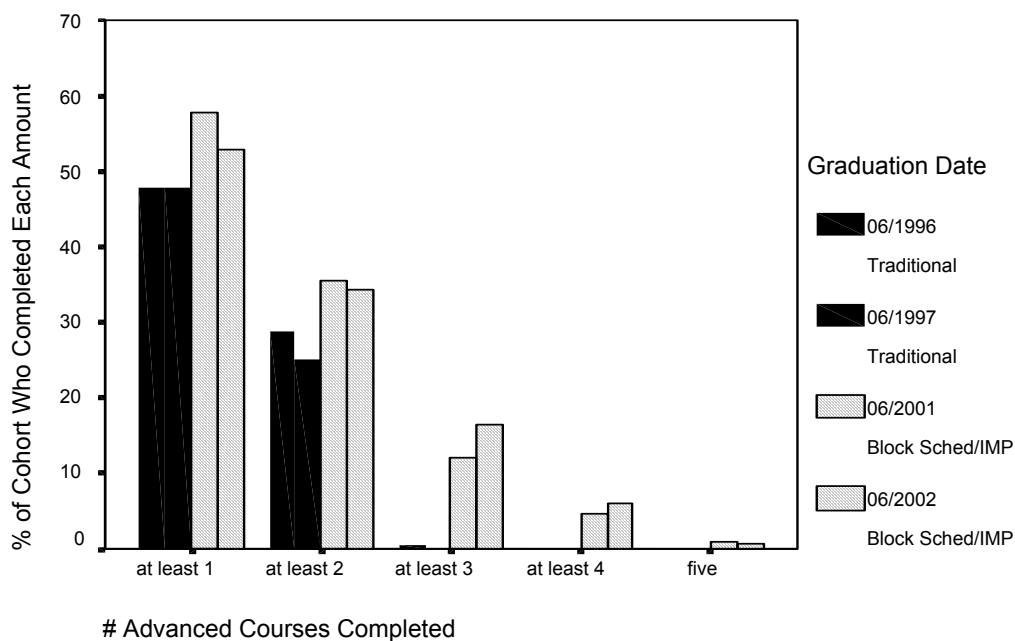
We also investigated the number of “Advanced” courses students completed. We defined an advanced course as one that required completion of the core mathematics curriculum as a prerequisite and included more advanced material than had been in the core curriculum courses. By these criteria, after completing Algebra 2 students in the two Traditional cohorts could enroll in any one of the following more advanced courses: Algebraic Analysis, Algebraic Analysis Honors, Pre-Calculus/Discrete Honors, Calculus AB AP, or Calculus BC AP. After completing Integrated Math 4, students in the Reform cohorts could enroll in any of the following more advanced courses: Algebraic Analysis, Discrete Analysis, Discrete Analysis Honors, Functional Analysis, Functional Analysis Honors, Statistics, Statistics AP, Calculus AB AP, or Calculus BC AP. Also, students in the graduating class of 2002 could enroll in a non-Advanced Placement calculus course, named “Intro to Calculus Honors,” which was added to the curriculum during their senior year.

Results

After Suburban High School adopted IMP as its core mathematics curriculum and introduced a semestered block schedule, enrollment in mathematics classes increased markedly. Over four years of high school, the average student in the traditional schedule/traditional curriculum graduating class of 1996 enrolled in 501.8 hours of mathematics classes and the average student in the traditional schedule/traditional curriculum graduating class of 1997 enrolled in 494.2 hours of mathematics classes. In contrast, students from the class of 2001, the first cohort who completed Grades 9 through 12 after the new curriculum and schedule were adopted, averaged 564.2 hours enrolled in mathematics classes. Students from the class of 2002, the second cohort who completed Grades 9 through 12 after the new schedule and curriculum were adopted, averaged 589.3 hours enrolled in mathematics classes.

The increased enrollment was particularly evident in advanced classes such as statistics, discrete analysis, functional analysis and calculus. Figure 28 shows the actual distribution of students enrolled in various numbers of advanced mathematics courses. As the figure shows, after the semestered block schedule and IMP curriculum were adopted, the percentage of students enrolled in at least one advanced course increased to include the majority of the student body. Also, under the new curriculum and schedule a noticeable number of students had the motivation and opportunity to enroll in three, four, or even five advanced courses.

Figure 28. Number of Advanced Course Completed.



Analysis of the Advanced Placement Program

Methodology

We utilized the automated transcript database to compute the number of students who received credit in Advanced Placement courses going back to the 1994/95 school year. In addition, Suburban High School publishes a yearly “School Profile” which reports various achievement benchmarks from the previous year, including the number of students who completed each Advanced Placement exam and their scores on those exams. We obtained information about Advanced Placement Exam grades from these School Profiles.

Results

Adoption of the semestered block schedule had a direct impact on the way Advanced Placement (AP) mathematics classes were taught. Previously, Suburban High School had offered students a choice of two AP mathematics classes: either a yearlong Calculus AB course

or a more rigorous year long Calculus BC course. Under the semestered block schedule, Suburban High School instead offered a one-semester Calculus AB course, followed by an optional second semester during which students completed Calculus BC.

The new schedule appeared to have a number of effects on student enrollment in Advanced Placement courses. While the number of students enrolling in calculus remained approximately the same as it had been under a traditional schedule, a larger percentage of students taking calculus completed BC Calculus in the two-semester format than had opted for the yearlong BC course under a traditional schedule. Also, under a traditional schedule calculus had been a course exclusively for seniors. Under the semestered block schedule it became common for top students to complete BC calculus in their junior year, followed by AP Statistics in their senior year. Meanwhile, large numbers of other students began taking the newly offered AP Statistics course, and by 2000/01 about half of those enrolled completed the AP statistics exam for college credit.

Figure 29 displays the way Advanced Placement course enrollment changed as the semestered block schedule was implemented. Note that in the “hybrid schedule” years 1997/98 and 1998/99 Advanced Placement students had utilized a traditional schedule for their early high school years but changed to a semestered block schedule when it was implemented school wide in 1997/98. In 2001/02 Suburban High School began offering a new non-AP calculus course named “Intro to Calculus Honors.” As the figure shows, this new course apparently siphoned off some of the students who would otherwise have enrolled in an Advanced Placement course.

Figure 29. Number of Students Passing Advanced Placement Courses.

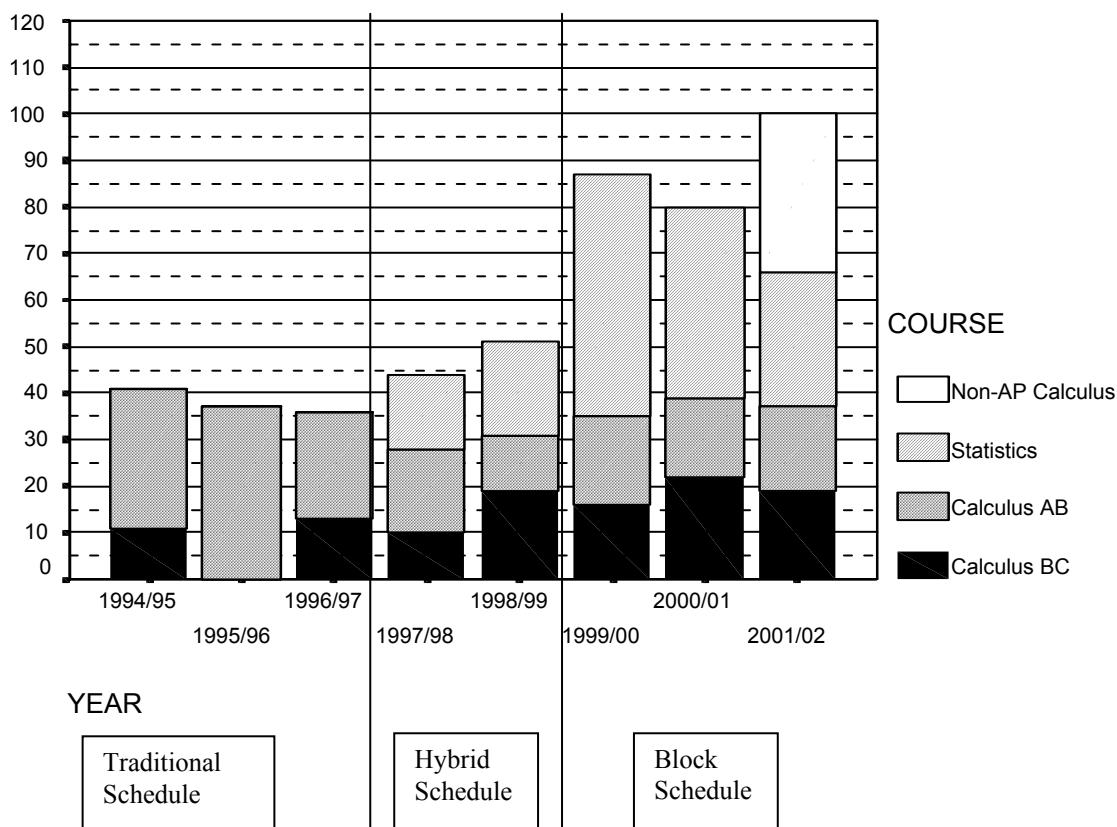


Table 5 displays Advanced Placement course enrollment and exam scores each school year from 1994-95 through 2001-02. To highlight the relationship between Advanced Placement performance and schedule/curriculum changes at Suburban High School, Table 5 uses the following conventions: Years with a traditional schedule and curriculum are displayed in standard type. Years with a semestered block schedule during which students taking AP courses had completed a traditional core curriculum are displayed in *italic* type. The one school year (1999-2000) with a semestered block schedule during which some students taking AP courses had completed a traditional core curriculum and others had used IMP is displayed in ***bold italic*** type. Years with a semestered block schedule during which students taking AP courses had all completed IMP are displayed in **bold** type.

In interpreting Table 5, the reader should note that prior to 1996-97 Suburban High School paid the fees for Advanced Placement Exams, and required all students enrolling in Advanced Placement courses to take the exam. This policy inflated the number of exam takers during the earlier years as shown in the table. For example, in 1994-95, the number of students taking the AB Calculus exam was larger than the number who received a passing grade in the course. Also, during 2000-01 and 2001-02 Advanced Placement Calculus AB was only offered during the fall semester. The AP exam is currently administered in the spring, and this delay appears to have discouraged students who completed the fall AB Calculus course but did not complete the second-semester BC course from taking the AB Calculus exam.

Table 5. Advanced Placement Enrollment and Exam Grades by School Year and AP

Mathematics Course

School Year			# Students Passing Course	Exam Grade					# Students Scoring 3 or Higher
				5	4	3	2	1	
1994-95	Course	Calculus AB AP	30	6	15	8	3	1	29
		Calculus BC AP	11	4	2	3	1	-	9
		Statistics AP	*	-	-	-	-	-	-
		Total	41	10	17	11	4	1	38
1995-96	Course	Calculus AB AP	37	5	11	13	2	3	29
		Calculus BC AP	**	-	-	-	-	-	-
		Statistics AP	*	-	-	-	-	-	-
		Total	37	5	11	13	2	3	29
1996-97	Course	Calculus AB AP	23	6	7	2	3	-	15
		Calculus BC AP	13	1	7	3	1	1	11
		Statistics AP	*	-	-	-	-	-	-
		Total	36	7	14	5	4	1	26
1997-98	Course	Calculus AB AP	18	1	7	2	1	-	10
		Calculus BC AP	10	-	2	-	-	-	2
		Statistics AP	16	2	4	-	-	-	6
		Total	44	3	13	2	1	-	18
1998-99	Course	Calculus AB AP	12	4	6	-	-	-	10
		Calculus BC AP	19	6	1	2	1	-	9
		Statistics AP	20	3	3	-	1	-	6
		Total	51	13	10	2	2	-	25
1999-00	Course	Calculus AB AP	19	2	2	1	-	-	5
		Calculus BC AP	16	12	2	-	-	-	14
		Statistics AP	52	9	6	-	-	-	15
		Total	87	23	10	1	-	-	34
2000-01	Course	Calculus AB AP	17	5	-	1	-	-	6
		Calculus BC AP	22	18	2	-	-	-	20
		Statistics AP	41	10	8	2	-	-	20
		Total	80	33	10	3	-	-	46
2001-02	Course	Calculus AB AP	18	1	-	-	-	-	1
		Calculus BC AP	19	15	2	2	-	-	19
		Statistics AP	29	11	2	-	-	-	16
		Total	66***	27	4	2	-	-	33

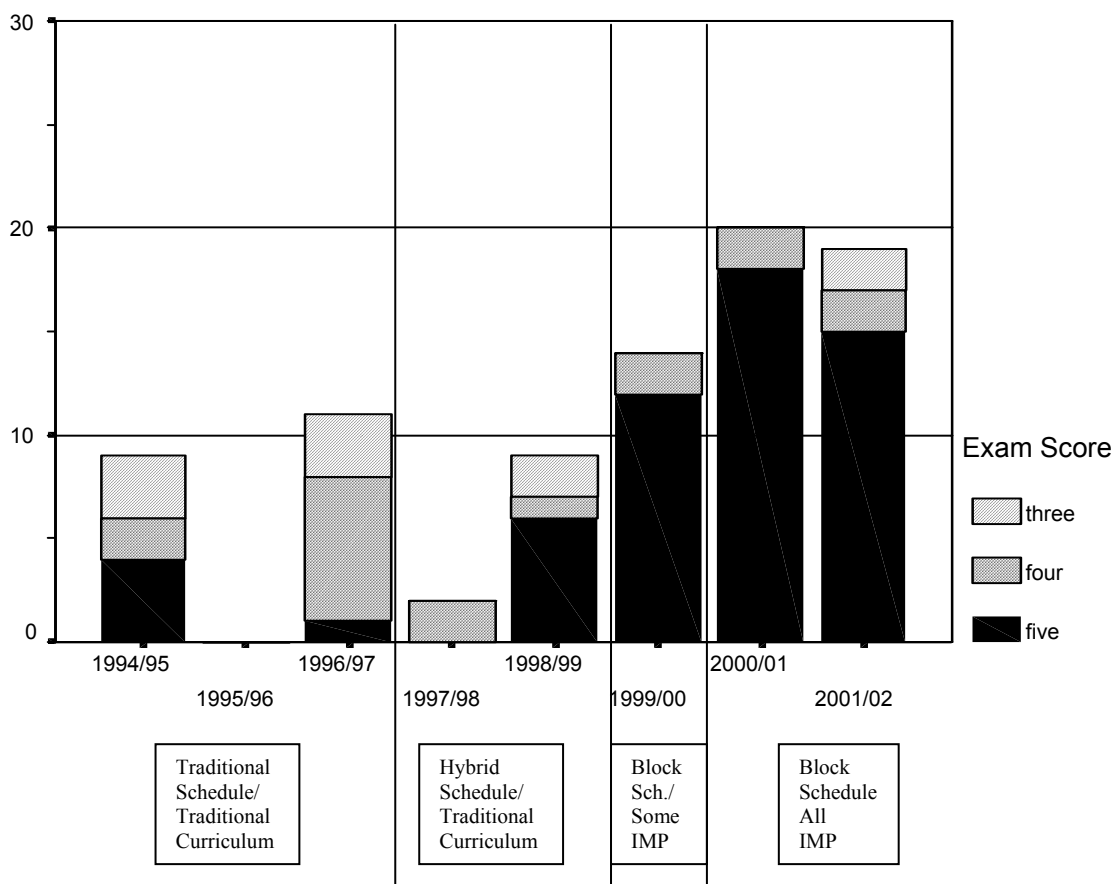
* Advanced Placement Statistics was not offered until 1997-98

** In 1995-96 too few students enrolled in Calculus BC for the course to be offered.

*** The total of 66 students completing AP courses in 2001-2002 does not include an additional 34 students who completed a newly offered non-AP course named "Intro to Calculus Honors."

As can be seen in Table 5, the block schedule by itself did not have any clear effect on AP exam grades. However, once the block schedule and IMP were combined, the number of students receiving the highest possible score of “5” (extremely well qualified) increased. This trend can be seen starting in 1999-00, when a significant fraction of AP students were juniors who had utilized the IMP curriculum, and became particularly pronounced in 2000-01, when all AP students had utilized IMP as their core curriculum. With a larger percentage of calculus students completing the more rigorous BC exam, and with generally higher AP exam grades, there was an impressive increase in the number of students achieving top marks on the BC calculus exam, as reflected in Figure 30.

Figure 30. Number of Students Completing BC Calculus Exam with a Passing Grade of “3” or Higher.



As shown in Figure 30, BC calculus was not offered in the traditional schedule/traditional curriculum year of 1995-96 because too few students enrolled in the course. In 1999/2000 the majority of students who completed BC Calculus were eleventh graders who had utilized IMP as their core curriculum, and the rest were seniors who had utilized a traditional core curriculum. In 2000/01 and 2001/02 all students enrolled in BC Calculus had utilized IMP as their core curriculum.

Discussion

Previous research had found a negative effect of semestered block scheduling on student mathematics achievement at schools where the mathematics curriculum was not adapted to fit the schedule. This study tested the theory that in contrast, simultaneous adoption of a semestered block schedule and a problem-centered curriculum, combined with appropriate planning and professional development, would result in improved mathematics achievement over what had been possible using a traditional schedule and curriculum.

Results of the study supported this theory. The profile of algebra achievement at the end of eleventh grade for students who used the semestered block schedule and IMP curriculum was different from that of students who used a traditional schedule and curriculum, but certainly not weaker. The block schedule/IMP students were better at formulating mathematical models, at interpreting graphs of complex functions, and at translating an exponential function from tabular form to an equation. They also appeared to be better at applying their algebra knowledge to solve a complex real-world problem: although the part the Algebra Achievement Test addressing this skill was completed by pairs and thus had a smaller sample size, the advantage of the Reform cohorts nonetheless approached statistical significance. Students using the traditional curriculum and schedule were better at simplifying expressions but only when those expressions used integer (not fractional) coefficients. Traditional cohort students were also better at

Draft

translating the graph of a line into an equation for that line. The two groups were approximately equal on other algebra skills tested, although there was a statistically non-significant trend for the Block Schedule/IMP group to do better when the skills were tested in context of an applied problem.

An important test of a core curriculum's success is the ability of students who complete it to pursue more advanced studies of mathematics. By this measure, utilizing IMP in the context of a semestered block schedule appeared to be very successful. Large numbers of students at Suburban High School used the opportunity provided by the new schedule and curriculum to study advanced mathematics. The proportion of students studying at least one advanced course increased, and many students enrolled in two or more such advanced courses.

By the end of their senior year, students in the Block Schedule/IMP cohorts were achieving better than students in the Traditional cohort in all areas of mathematics tested. This improvement included the content areas of "Data Analysis, Statistics, and Probability," "Geometry and Measurement," and "Algebra and Number."

While results of the Algebra Achievement Test administered at the end of eleventh grade reflected positively on the combination of semestered block schedule and IMP curriculum, it is clear that results of the NAEP blocks administered at the end of twelfth grade were still more positive. We offer two possible explanations for this improved performance. One possible explanation is that when the Algebra Achievement Test was administered at the end of eleventh grade, a large percentage of students in the Block Schedule/IMP cohorts had not yet completed the four IMP courses in the core curriculum. In the graduating class of 2001 (the First Reform cohort), 42.3% of the student body completed Integrated Math 4 in their senior year. In the graduating class of 2002 (the Second Reform cohort), a smaller but still sizeable 29.7% of the student body completed Integrated Math 4 during their senior year. In contrast, in the graduating

Draft

class of 1998 (the Traditional cohort who completed the Algebra Achievement Test and the NAEP blocks), only 4.8% were enrolled in Algebra 2 during their senior year. An additional 12.4% were enrolled in Algebra 3/Trig their senior year, but unlike students in the Reform cohorts who had not completed Integrated Math 4, the students in Algebra 3/Trig had already completed some version of their core curriculum by the time they were administered the Algebra Achievement Test at the end of eleventh grade. (Algebra 3/Trig was designed for students who had received low grades in Algebra 2 or who had completed a less advanced course named Algebra 2 Career/College Prep instead of the regular Algebra 2 course. It covered Algebra 2 from a different perspective, using the text *Advanced Algebra* (Senk & Thompson, 1993).)

A second possible explanation for the greater contrast between block schedule/IMP students and the traditional schedule/traditional curriculum students on the NAEP blocks than on the Algebra Achievement Test concerns the nature of the test items. Recall that on the Algebra Achievement Test, Traditional cohort students scored higher on items testing routine procedures presented in well-practiced formats, and students in the Reform cohorts scored higher on nearly all other types of items. In the three released NAEP blocks, there were no items that could be considered purely “routine.” Even those items categorized as “procedural” required some degree of problem solving or conceptual understanding. A sample item categorized by NAEP as “procedural” is Item 2 from NAEP Block P3 (1996-12M13), the only NAEP item on which Reform cohort students did not outscore Traditional cohort students. That item instructed students: “In the space below, use your ruler to draw a parallelogram that has perpendicular diagonals. Show the diagonals in your sketch.”

While most analyses reported in this study contrasted achievement of students utilizing the IMP curriculum combined with a semestered block schedule to that of students utilizing a traditional core curriculum combined with a traditional schedule, for Honors students we were

Draft

able to make an additional comparison. We were able to contrast the achievement of Honors students who utilized the IMP curriculum combined with a semestered block schedule to that of Honors students who used a traditional core curriculum combined with a semestered block schedule. Results indicated that Honors students who had utilized IMP were superior in Conceptual Understanding, with a substantively and statistically significant advantage of 0.57 standard deviations on NAEP conceptual understanding items.

After IMP was adopted as a core curriculum, there was a marked increase in the total number of students enrolled in Advanced Placement mathematics courses, including an increase in the number of students who completed the exam in BC calculus, the most demanding mathematics course traditionally offered at the high school level. Further, BC calculus exam grades improved to an impressive level. In the spring of 2001, the first year when all students enrolled in AP Calculus had utilized IMP as their core curriculum, of twenty students who completed the BC exam, two received a score of “4” (well qualified), and the remaining eighteen received the highest possible score of “5” (extremely well qualified), for a mean score of 4.9. That same year, there were only seven other public high schools in the United States where at least twenty students completed the exam and the mean score was 4.9 or higher. In the spring of 2002 the number of students receiving top marks in the BC calculus exam, although retreating slightly from the stellar numbers of the previous year, nonetheless remained considerably higher than had been the case under the traditional curriculum and schedule.

This current study was not a controlled experiment investigating success in Advanced Placement Calculus. Several things changed in addition to curriculum and schedule, including both the Calculus teacher and the Calculus textbook. Furthermore, because the semestered block schedule offered students the opportunity to take more courses, teachers at Suburban High School decided to have their top students complete the four IMP texts, followed by a one

Draft

semester functional analysis course, followed by Advanced Placement Calculus. Thus, the study cannot be viewed as a test of the IMP curriculum authors' belief that IMP by itself is sufficient preparation for studying calculus (Fendel, Resek, Fraser, & Alper, 1997; Green, 2000).

Nonetheless, this study can be viewed as a strong “existence proof” that IMP can in fact be a very successful core curriculum for students who wish to pursue advanced mathematics. These results counter widely publicized concerns that had been expressed by critics of the IMP curriculum (An Open Letter to United States Secretary of Education, Richard Riley, 1999; Wu, 2000).

Overall, it appears clear that combining a semestered block schedule with the problem-centered IMP curriculum succeeded in improving student mathematics achievement at Suburban High School. This finding is particularly important in view of past research indicating that mathematics achievement is likely to suffer at schools that adopt a semestered block schedule without modifying their mathematics curriculum.

We have called this study “a tale of synergy.” It is our belief that the positive results we reported were due to a schedule innovation and a curriculum innovation working in concert. Both our literature review and our study of Honors students in the Pilot cohort provided evidence that such a positive result would not have been achieved by a semestered block schedule implemented alone, without a change in curriculum. However, our evidence does not address whether or not similarly positive results could have been achieved by adopting the IMP curriculum alone, without a semestered block schedule. It seems plausible that results would have been less positive without the extra time for studying mathematics available to students under the semestered block schedule, but we cannot be certain.

Limitations and Suggestions for Further Research

A primary limitation of this study is the unusual nature of the Suburban High School community. An unusually large percentage of parents in the community were highly educated professionals, and the typical student entering Suburban High School had prior achievement well above the national average. The nature of the local community probably affected the results of this study, for example encouraging students to take more mathematics courses once the semestered block schedule gave them the opportunity to do so.

Also, most students who participated in this study both in the Traditional cohort and in the Block Schedule/IMP cohorts attended Suburban Middle School, which had implemented a mathematics curriculum viewed as being particularly compatible with IMP. The positive effects of the High School reform program may have been enhanced by the students' middle school experience.

Finally, the Mathematics Department at Suburban High School provided unusually fertile ground where an innovation like the one described here might grow. They helped plan the semestered block schedule. They sought out the IMP curriculum, choosing it as the best match available for a set of objectives they had already developed. They participated in many hours of professional development. There was low teacher turnover, with many mathematics teachers remaining at the High School from before the intervention was initiated in 1996/97 through the present time.

Given the unique characteristics of Suburban High School, it is important to replicate the findings of this study at other sites with differing community and school profiles. Other sites attempting to replicate the program at Suburban High School may not have as compatible a middle school curriculum, as low a rate of teacher turnover, or a faculty with prior experiences as well matched to the IMP curriculum as was the case in this study. Nonetheless, we are optimistic that the findings will be positive. Previous literature indicates that, taken

Draft

individually, each of the two innovations that were combined to comprise the successful reform at Suburban High School might produce even more positive effects at a school with a lower profile of prior achievement. Many of the potential advantages of a semestered block schedule, including a calmer school atmosphere, fewer discipline referrals, and fewer drop-outs, are likely to be more important at a school with a lower prior achievement profile than was the case at Suburban High School (Kramer, 1997a). Further, even critics of the IMP curriculum have noted that IMP is likely to be particularly effective for students who have historically been disaffected from mathematics (Wu, 2000).

We view the current study as a first test of the theory that a semestered block schedule, combined with a problem-centered curriculum designed to implement the principles of the NCTM *Standards* documents (NCTM 1989, 1991, 2000), is likely to improve the mathematics learning of high school students. These predicted results contrast with the effects of a semestered block schedule implemented with an unchanged mathematics curriculum, a combination that previous research indicates can sometimes lead to lower mathematics achievement. We believe the combination of a semestered block schedule and an NCTM *Standards*-based curriculum has passed the first test. We look forward to learning the results of other tests conducted at other sites.

References

- Allison, Paul D. (2002). Missing data. Thousand Oaks, CA: SAGE.
- Averett, Chris P. (1994). Block scheduling in North Carolina high schools. Raleigh, NC: North Carolina Department of Public Instruction.
- Bourgeacq, P. B., Brungardt, C. B. Forinash, L., Helscher, M. L., Kennel, T., Reynolds, L. L., Robinson, T., Schulte, M., Smith, C. Stearns, P. M., & Thayer, B. J. (1997). Processing and scoring assessment materials. In N. L. Allen, F. Jenkins, E. Kulick, and C. A. Zelenak (Eds.) Technical Report of the NAEP 1996 State Assessment Program in Mathematics. Washington, DC: National Center for Education Statistics.
- Carroll, John M. (1994). The Copernican plan evaluated: The evolution of a revolution. Topsfield MA: Copernican Associates.
- Fendel, Dan, Resek, Diane Fraser, Sherry & Alper, Lynne (1997). What is the place of algebra in the K-12 mathematics curriculum? Bulletin: National Association of Secondary School Principals, 81(586), 60-63.
- Foreman, Linda C. & Bennet, Albert B. (1991) Visual Mathematics, Course I-III. Salem, OR: Math Learning Center.
- Green, Lori (2000). Teaching handbook for the Interactive Mathematics Program. Retrieved April 27, 2002, from <http://www.mathimp.org/resources/teacher/index.html>
- Hackman, D. G. (1995). Ten guidelines for implementing block scheduling. Educational Leadership 53(3), 24-27.
- Harter, Randy (1994). The pros and cons of alternative high school schedules. In C. Averett (Ed.) Block Scheduling in North Carolina High Schools. Raleigh, NC: North Carolina Department of Public Instruction.
- Huntley, M. A., Rasmussen, C. L., Villarubi, R. S., Sangtong, J. & Fey, J. T. (2000) Effects of
- Draft

- standards-based mathematics education: A study of the Core-Plus Mathematics Project algebra and functions strand. Journal for Research in Mathematics Education, 31(3), 328-361.
- Klein, David, Askey, Richard, Milgram, R. James, Wu, Hung-Hsi, Scharlemann, Martin, & Tsang, Betty (1999). An Open Letter to United States Secretary of Education, Richard Riley. Retrieved April 28, 2002, from <http://www.mathematicallycorrect.com/riley.htm>.
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. International Journal of Artificial Intelligence in Education, 8, 30-43.
- Kramer, Steven L. (1996). Block scheduling and high school mathematics instruction. Mathematics Teacher, 89 (9), 758-768.
- Kramer, Steven L. (1997a, February). What we know about block scheduling and its effects on math instruction, Part 1. Bulletin: National Association of Secondary School Principals, 81(586), 18-42.
- Kramer, Steven L. (1997b, March). What we know about block scheduling and its effects on math instruction, Part 2. Bulletin: National Association of Secondary School Principals, 81(587) 60-82.
- Kramer, Steven L. (2001). The joint impact of block scheduling and a standards-based curriculum on high school algebra achievement and mathematics course taking. Unpublished doctoral dissertation, University of Maryland, College Park. Available online at <http://www.gphillymath.org/BlockSchedule/>
- Marshall, M., Taylor, A., Bateson, D., & Brigden, S. (1995). The British Columbia assessment of mathematics and science: Preliminary report (DRAFT). Victoria, B.C.: B.C. Ministry of Education.

- Meadows, Martha E. (1995). A preliminary program review of the four-period day as implemented in four high schools. Unpublished doctoral dissertation, University of Maryland, College Park.
- Merlino, F. Joseph, Wolff, Edward, & Tolbert, Clara (2000). Assessing the cost/benefits of an NSF “Standards-based” secondary mathematics curriculum on student achievement: The Philadelphia Experience. Presentation before the National Science Foundation. Philadelphia, PA: Greater Philadelphia Secondary Mathematics Project.
- National Council of Teachers of Mathematics (1980). An agenda for action: Recommendations for school mathematics of the 1980s. Reston, VA: National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics (1989). Curriculum and Evaluation Standards for School Mathematics. Reston, VA: National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics (1991). Professional standards for teaching mathematics. Reston, VA: National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics (1993). Assessment standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics (2000). Principles and Standards for School Mathematics. Reston, VA: National Council of Teachers of Mathematics.
- North Carolina Department of Public Instruction (1997). Block scheduling in North Carolina: Implementation, Teaching, and Impact Issues 1997 Survey Results. Retrieved April 25, 2002, from <http://www.dpi.state.nc.us/reportstats.html>.
- Raphael, Dennis, Wahlstrom, Merlin, & McLean, L.D. (1986). Debunking the semestering myth. Canadian Journal of Education, 11(1), 36-52.
- Reid, W. M. (1995a). Restructuring secondary schools with extended time blocks and intensive
Draft

- courses: The experiences of school administrators in British Columbia. Unpublished doctoral dissertation, Gonzaga University, British Columbia.
- Reid, W. M. (1995b). Restructuring a Secondary School with Extended Time Blocks and Intensive Learning. Nelson, British Columbia: unpublished manuscript.
- Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. Journal of Educational Measurement, 29, 133-61.
- Rubin, Donald B. (1987). Multiple imputation for nonresponse in surveys. New York: Wiley.
- SAS Institute (2001). SAS/Stat Software: Changes and enhancements, release 8.2. Cary, North Carolina: SAS Institute. Retrieved October 10, 2002 from <http://support.sas.com/rnd/app/da/new/802ce/stat/index.htm>.
- Senk, S. L., Thompson, D. R., Viktora, S. S., Usiskin, Z., Ahbel, N. P., Weihold, M. L., Rubenstein, R. N., Jaskowiak, J. H., Flanders, Jakucyn, N. & Pillsbury, G. (1993). Advanced algebra (2nd ed. field-trial version). Chicago: University of Chicago School Mathematics Project.
- Sessoms, J. C. (1995). Teachers' perceptions of three models of high school block scheduling. Unpublished doctoral dissertation, University of Virginia, Charlottesville.
- Stevens, R. S. (1976). Semestering in Vancouver schools. Vancouver, British Columbia: Vancouver Board of School Trustees. (Eric Document Reproduction Service No. ED 131 556)
- Webb, Norman L. (in press). The impact of the Interactive Mathematics Program on student learning. In S. Senk and D. Thompson (Eds.) Standards-oriented school mathematics curricula: What does the research say about student outcomes? Mahwah, NJ: Erlbaum.

Wild, Reginald (1998, April). Science Achievement and Block Schedules. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Diego, California.

Zhang, Gongshu (2001, April). Academic Performance Differences between Students in Block and Traditionally Scheduled High Schools 1993-2000. Paper presented at the annual meeting of the American Educational Research Association, Seattle, Washington.