The Joint Impact of Block Scheduling and an NCTM Standards-based Curriculum on High School Mathematics Achievement

By Dr. Steven Kramer, University of Maryland
March 5, 2003

Summary

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.

Over 4 years of high school, students using the block schedule and IMP on average enrolled in more hours of mathematics courses than had earlier groups of students who used a traditional schedule and curriculum. Under the block schedule, sizeable numbers of students completed the four IMP courses and went on to take subsequent advanced mathematics courses. They were more likely to take one, two, three, four, or even five advanced mathematics courses than had been the case under the traditional schedule and curriculum, despite the fact that under IMP they were required to complete four core courses before doing so, whereas students using the traditional curriculum and schedule had been required to complete only three core courses before moving on.

At the end of Grade 11, students using the block schedule and IMP were better able to formulate algebraic models, interpret graphs and tables, solve algebra problems presented in context, and work in pairs to solve an extended, open-ended, applied algebra problem. Students who had used a traditional schedule and curriculum were better able to translate a line graph into an equation and to simplify expressions with integer (but not fractional) coefficients.

By the end of Grade 12, students in the Block Schedule/IMP cohorts were achieving better than students in the Traditional cohort in all areas of mathematics tested, outscoring Traditional cohort students on 27 of 28 publicly released items from the 1996 National Assessment of Educational Progress (NAEP). These items covered the content areas of "Data Analysis, Statistics, and Probability," " Geometry and Measurement," and "Algebra and Number." Within each content area, students in the Block Schedule/IMP cohorts achieved higher than those in the Traditional cohort on Problem Solving, Conceptual Understanding, and Procedural items.

A group of 34 Honors students who utilized a traditional algebra 1, geometry, algebra 2 curriculum during the first years when the block schedule was implemented enabled us to compare their achievement with later high ability students who used IMP as their core curriculum under a semestered block schedule. The IMP groups scored significantly and

substantively (0.56 standard deviations) higher on NAEP Conceptual Understanding items than traditionally taught students.

Under the semestered block schedule about 13% of the student body enrolled in a newly offered Advanced Placement (AP) statistics course. AP Calculus enrollment remained about constant after the school adopted the block schedule and IMP, but a larger percentage of AP Calculus students completed the more demanding BC course, and Advanced Placement exam grades improved. In the first cohort of students to utilize both the semestered block schedule and the IMP curriculum, twenty students completed the BC calculus exam. Eighteen of those twenty students received a score of "5" and two received a score of "4" on the test, for a mean score of 4.9. That year, there were only seven other high schools in the United States where at least twenty students completed the BC calculus exam and the school-wide mean score was 4.9 or higher.

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.

During the 1990s, researchers developed several new high school curricula specifically designed to conform to the vision described in the *Standards*. There has been considerable controversy about the worth of these new curricula ("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 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 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. Block schedules have been used widely in parts of Canada since the 1970s, well before they became popular in the United States. A number of large-scale studies have investigated achievement effects of block scheduling in Canada (Raphael, Wahlstrom, & McLean, 1986; Marshall, Taylor, Bateson, & Brigden, 1995; Wild, 1998). They have consistently reported that mathematics achievement under a semestered block schedule is lower than under a traditional schedule.

In contrast, anecdotal reports (Kramer, 1996), as well as some achievement data from North Carolina (Averett, 1994; North Carolina Department of Public Instruction, 1997) 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 larger number of courses. In the United States, students generally take a larger number of 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 larger number of courses (Harter, 1994; Kramer, 1996). Unfortunately, in most cases research on achievement effects of block scheduling has been conducted at sites where this approach has not been implemented. 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 School[1]

This research project is part of an ongoing research effort that was begun during the 1996-97 school year by teachers at Suburban High School. Suburban is the only high school in a small school district located in a middle class suburb of a large U.S. city. 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 competed 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. 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, thus creating the conditions previous research theorized might improve mathematics achievement under a block schedule (Kramer, 1996).

Grade 11 Algebra Achievement Test

Testing the Traditional cohort. In order to assess their understanding of algebra, in the spring of 1997 all eleventh graders at Suburban High School were administered an Algebra Achievement test. Eleventh graders tested in 1997 comprise a “Traditional” cohort who used a traditional schedule and traditional curriculum throughout Grades nine through eleven.

The Algebra Achievement 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. Test 1 focuses on students’ ability to understand and solve algebra problems presented in context, as is typically emphasized by IMP and other reform curricula. Test 2 focuses on context-free symbolic manipulations of a type typically emphasized by traditional algebra texts. Items in Test 2 were adapted from released ACT examinations and from items that commonly appeared on college placement tests. Test 3 requires collaborative work on a single extensive open-ended applied problem and is completed by students working in pairs.

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 "Reform" 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. The IMP curriculum spends more time addressing Probability and Statistics than does the traditional algebra 1/geometry/algebra 2 sequence. In the area of geometry, the IMP curriculum probably allocates somewhat less time to geometric topics than does a traditional sequence including a full year of Euclidean geometry, but IMP addresses geometry in a more applications-oriented manner that the curriculum authors believe will enable students to be more successful at learning geometry concepts. To assess student achievement these content areas, teachers at Suburban High School decided to administer the three publicly released blocks of items from the 1996 twelfth grade mathematics assessment administered by the National Assessment of Educational Progress (NAEP). Approximately two-thirds of the items on the NAEP blocks addressed Data Analysis/Probability/Statistics or Geometry/Measurement. 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" "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 at Suburban High School completed the three NAEP blocks. Unfortunately, the research team did not obtain the NAEP blocks until the 1997-98 school year, after the entire school had switched to a semestered block schedule. Twelfth graders enrolled at Suburban High School at that time had utilized a traditional core curriculum. They had been enrolled in a traditional schedule throughout grades 9-11, but in a semestered block schedule in grade 12. 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 comprise a "Traditional Grade 12" cohort who used a traditional curriculum throughout Grades nine through twelve, and a traditional schedule throughout Grades nine through eleven.

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 Grade 12" cohort. They were the first group to utilize a semestered block schedule throughout Grades 9-12. Further, non-Honors students in this group were the first to utilize IMP as their core curriculum. Teachers at Suburban High School were novices at using IMP's unusual approach to instruction, and it was anticipated that results obtained from using IMP with Pilot cohort students might be less positive than would be obtained with subsequent cohorts, after teachers had time to familiarize themselves with the IMP curriculum.

Finally, Honors students in the Pilot cohort utilized a semestered block schedule, but studied the traditional core mathematics sequence of algebra 1/geometry/algebra 2. Contrasting achievement of Honors students from the Pilot cohort to that of Honors students in subsequent

Reform cohorts will provide particularly useful information, as it will enable 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 at Suburban High School who were enrolled in mathematics class during the second semester of their senior year again completed the three NAEP blocks. Twelfth graders tested at these times comprise the "Grade 12 Reform" cohorts. Both Honors and non-Honors students 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 test scores and NAEP block test scores, students in the Reform 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.

Statistical Method

This study addressed three separate questions regarding student performance on mathematics achievement tests:

1) Was there a difference in algebra achievement by the end of eleventh grade between students in the Block Schedule/IMP cohorts and students in the Traditional cohort?

2) Was there a difference in achievement on blocks of NAEP items administered at the end of twelfth grade between students in the Block Schedule/IMP cohorts and students in the Traditional cohort?

3) Was there a difference in achievement on blocks of NAEP items administered at the end of twelfth grade between Honors students in the Pilot cohort, who utilized a

traditional core mathematics curriculum under a semestered block schedule, and Honors students in the Block Schedule/IMP cohorts, who utilized IMP as their core curriculum under a semestered block schedule?

We looked at multiple subscales to address each of these questions. Each of the three questions addressed a separate issue, which by itself could have merited a research article. Further, we used a slightly different statistical approach for each of the three questions, due to differences in the nature of the dependant variable and the availability of covariates. For these reasons, later in this article we report details of the statistical methodology we used separately for each of the three achievement questions. We used a Bonferonni adjustment within each question to maintain an overall within-question confidence level of .05.

In addition to addressing the above questions about the mathematics achievement tests we administered, we looked at three other measures of student learning: overall time enrolled in mathematics classes; number of advanced mathematics courses completed; and participation and achievement in Advanced Placement mathematics classes. Transcripts and information about Advanced Placement achievement were available for a number of cohorts previous to the Traditional group who completed the Algebra Achievement test and the NAEP blocks, so for these analyses we were able to use several earlier years' data to establish a firmer baseline. While it would have been possible to perform statistical analyses about changes in course enrollment, given the large sample size even a small change would have appeared statistically significant, so we did not think such an analysis would be useful. Also, because changes in Advanced Placement course offerings and enrollment over time were rather complex, we decided that a descriptive and graphical analysis of these changes would prove more enlightening than would a statistical hypothesis test.

Overview of Results

This study compared Reform cohort students to Traditional cohort students in four areas: algebra knowledge at the end of eleventh grade, mathematics course enrollment, achievement on the NAEP blocks at the end of grade 12, and Advanced Placement participation and exam scores. At the end of eleventh grade, while the block schedule/IMP group did better on most algebra skills measured, there were other algebra skills on which the traditional group was superior. By the end of twelfth grade, achievement of the block schedule/IMP group was superior to that of the traditional group in all areas measured, with students in the Reform cohorts scoring higher than students in the Traditional cohort on 27 of the 28 NAEP items administered.

Students who completed the IMP curriculum under a semestered block schedule were more likely than students who completed a traditional core curriculum under a traditional schedule to enroll in advanced mathematics classes such as discrete analysis, statistics, functional analysis, and calculus. Also, the Advanced Placement program appeared to be strengthened under the combination of an IMP curriculum and a semestered block schedule. A larger number of students enrolled in Advanced Placement courses. Further, Advanced Placement exam grades improved markedly.

This study also found that Honors students who utilized the IMP curriculum under a semestered block schedule were better at NAEP items measuring conceptual understanding than were Honors students who utilized a Traditional curriculum and under a semestered block schedule.

Grade 11 Algebra Achievement Test: Methodology

Our analysis of the Algebra Achievement Test administered at the end of eleventh grade was based on a dissertation completed by Kramer (2002), but improved on the earlier work in a number of respects. First, the primary statistical analysis reported by Kramer analyzed each part of the Algebra Achievement test as a separate subscale. While providing a good overall picture of student achievement, this decision obscured important differences between the Block Schedule/IMP cohorts and the Traditional cohort on particular skills tested within Part 1 and Part 2 of the test. The current study follows Kramer in using a Bonferonni adjustment to allocate equal weight to each of the three parts of the test (i.e. using a confidence level of .017 for Part 3 and a confidence level of .033 for Parts 1 and 2 together). However, it looks at fifteen subscales within Parts 1 and 2, using a Bonferonni-adjusted confidence level of .033/15=.0022 for each.

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 instead of the continuous normally distributed data needed for Ordinary Least Squares Regression and related statistical techniques. For this reason, we analyzed the each dependant variable with a Probit analysis, with Grade 6 ability used as a covariate.

A second improvement on the earlier analysis was our redefinition of the way we measured Grade 6 ability. Kramer (2002) had utilized a combination of four variables; in contrast, the current analysis uses a single variable to measure Grade 6 ability. This single variable is the average of each student's sixth grade score on "quantitative ability" subscale of the Comprehensive Testing Program (CTP) test published by the Educational Records Bureau, and his or her 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. The single variable has high face validity, is more parsimonious than the collection of four variables, and explains as much of the variance in the dependant variables as did all four together.

In a final improvement, the current analysis uses multiple imputation (Rubin, 1987; Mislevy, Beaton, and Sheahan, 1992; Allison, 2002) to deal with missing data. The basic idea of multiple imputation is this: Given missing data, the best solution 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 randomly select missing values from the probability distribution of what those values are likely to be, and run the analysis with these randomly-selected values. Repeat this process several times, and use as your result the average of the several runs. Each set of imputed values randomly selected from the probability distribution of possible values 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.

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. Thus, when a student's ability score was missing we imputed fifteen plausible values for the missing score, using a Markov Chain Monte Carlo process from SAS Proc MI (SAS Institute, 2002a). As recommended by Allison (2002) for 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. 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 and scores from the Reform 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 reported as an effect size the average effect size across the fifteen runs. We used SAS Proc Mianalyze (SAS Institute, 2002b) 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.

Multiple Imputation had the greatest impact on analysis of Part 3. Because Part 3 was completed by students working in pairs there were two ability scores for each Part 3 test, leaving a high likelihood that one or the other would be missing. Kramer (2002) employed less current statistical techniques, requiring a choice between deleting a very large number of student scores that had at least one ability score missing, or else utilizing the average of the two ability scores when both were available, and utilizing the one available ability score when the other was missing. Neither was a particularly good choice. As explained by Rubin (1987) and Allison (2002) multiple imputation provides a superior and statistically rigorous alternative. For Part 3, the average of the two ability scores was found to be as good a predictor of a pair's achievement as was entering the two scores separately, so the single average score was used as a covariate.

Because we wanted to compare the overall effects during Grades 9-11 or 9-12 of a Traditional Schedule/Traditional Core Curriculum to those of a Semestered Block Schedule/IMP Core Curriculum, we deleted from the analysis the scores of any students who had not attended

Suburban High School beginning in 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.

Grade 11 Algebra Achievement Test: Results. Results on Algebra Achievement Tests 1 and 2 are best described by grouping the items into the nine sub-skills described in Table 1.

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

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 a mathematical model (from Test 1);
5. evaluating an expression when it is
   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
   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)

After controlling for prior ability, 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[2]. They were better at formulating a mathematical model to describe a problem situation ($p<.0004$; effect size = 0.32 standard deviations; 99.78% confidence interval for effect size = (.04,.60)). See Figure 1 for a sample item from the "formulating models" subscale and for a display of the raw (unadjusted for prior ability) difference between groups in percent of maximum possible score on this subscale.

Block Schedule/IMP students were also better at interpreting graphs on a Cartesian plane ($p<.0001$; effect size = 0.48 standard deviations; 99.78% confidence interval for effect size = (.17,.79)). See Figure 3 for a description of items on the "Interpreting Graphs" subscale and Figure 4 for a display of the raw (unadjusted for prior ability) difference between groups in percent of maximum possible score on this subscale. Finally, students in the Block Schedule/IMP cohorts were significantly better at translating an exponential function from tabular form to an equation (($p=.0002$; effect size = 0.53 standard deviations; 99.78% confidence interval for effect size = (.10,.96)). See Figure 5 for a description of the item measuring this ability, and Figure 6 for a display of the raw (unadjusted for prior ability) difference between groups in percent of maximum possible score on this subscale.

[2] All effect sizes and significance tests reported in this section are based on probit analyses after controlling for prior ability and performing multiple imputation. However, a supplemental analysis performed without controlling for prior ability identified the same subscales as statistically significant.

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

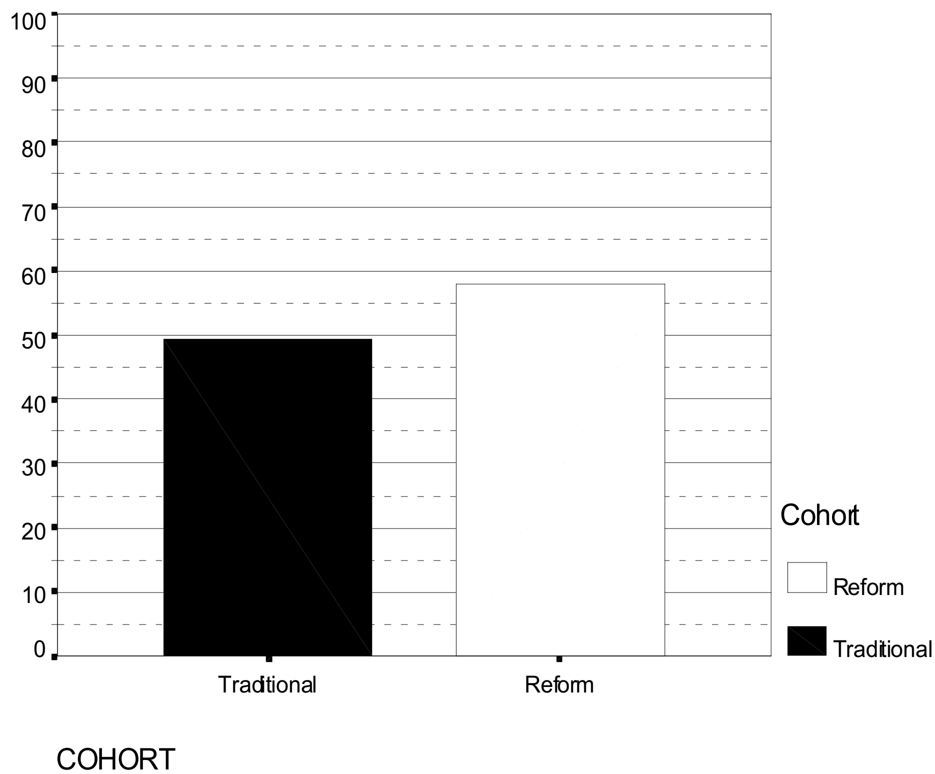
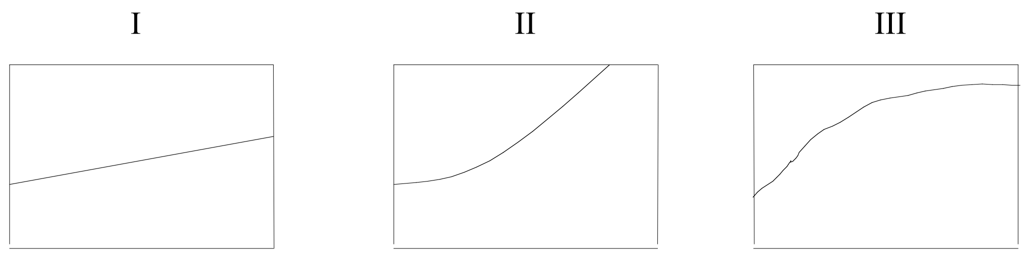


Figure 3. Questions Requiring Students to Interpret a Graphs on the Cartesian Plane

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.



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

a) at an increasing rate? _______
b) at a constant rate? _______
c) 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?

a) $S = 500 + 60t$ _______
b) $S = -15t^2 + 290t + 500$ _______
c) $S= 500 (1.2)^t$ _______

Figure 4. Mean Score Interpreting Graphs on the Cartesian Plane

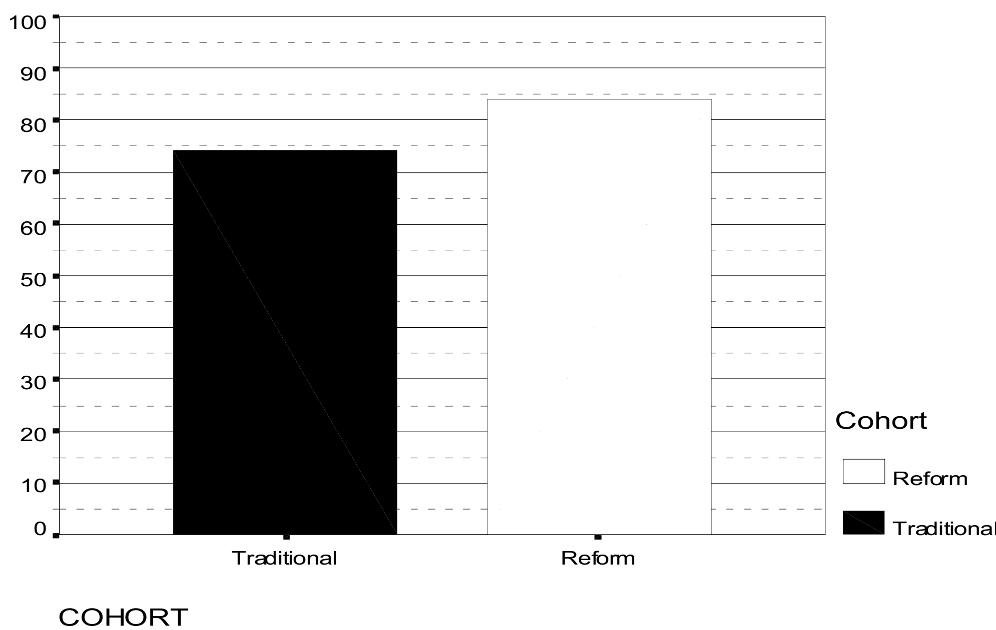
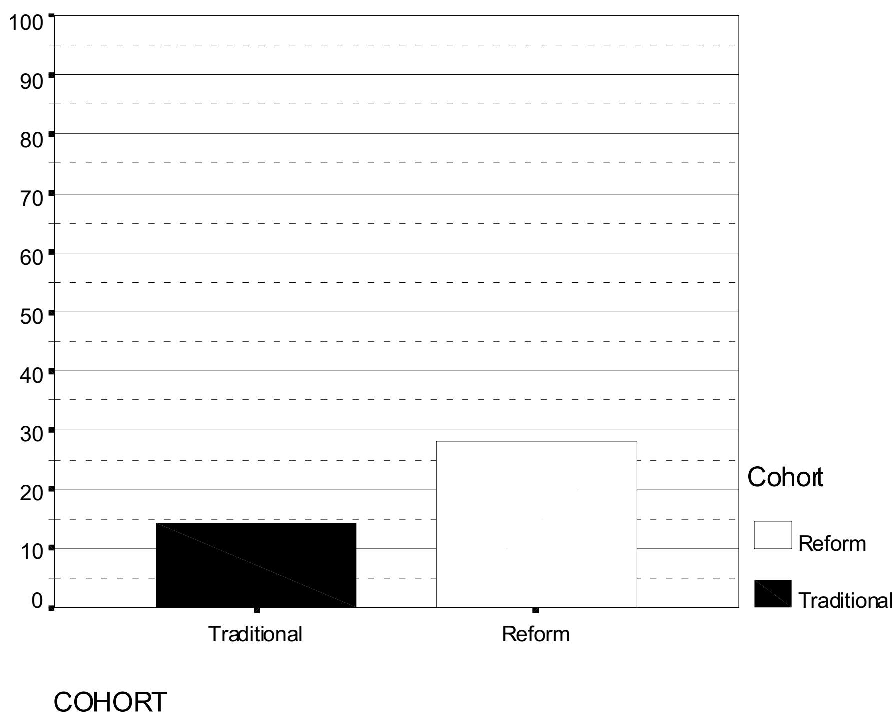


Figure 5. Question Requiring Students to Translate an Exponential Function from Tabular Form to an Equation

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

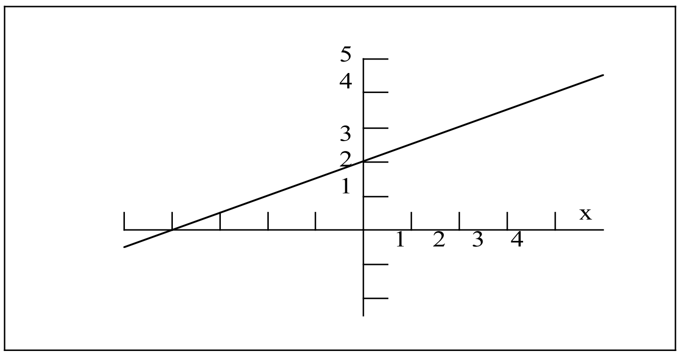
Figure 6. Mean Score Translating an Exponential Function from a Table to an Equation



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 ($p<.0001$; effect size = -0.48 standard deviations; 99.78% confidence interval for effect size = (-.79,-.17)). Figure 7 displays the items comprising this subscale and Figure 8 displays the raw (unadjusted for prior ability) difference between groups in percent of maximum possible score on this subscale. Students in the Traditional cohort also scored significantly higher at simplifying expressions when the expressions had integer coefficients (p=.0002; effect size = -0.48 standard deviations; 99.78% confidence interval for effect size = (-.88,-.09))--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?

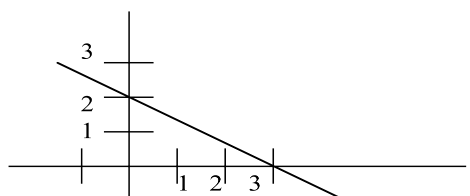


Figure 8. Mean Score Translating a Line Graph to an Equation

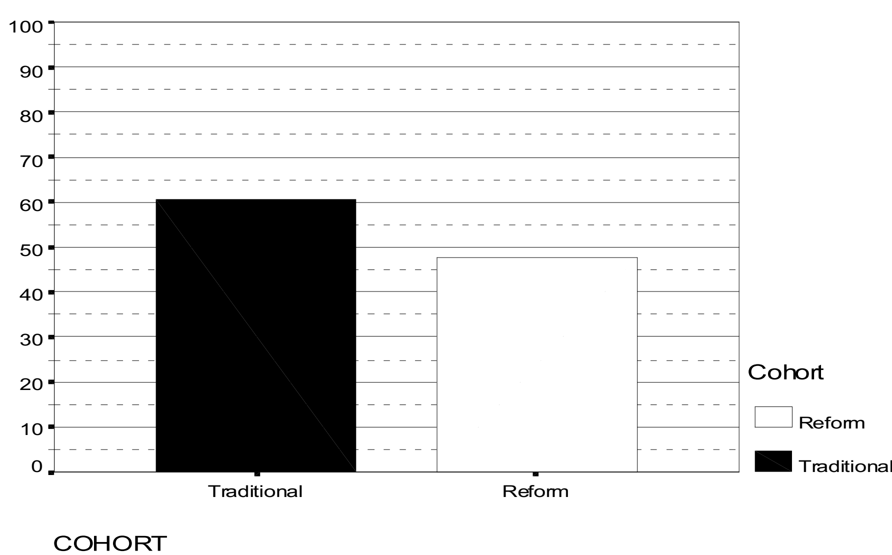


Figure 9. Stems from Multiple-Choice Questions Requiring Students to Simplify Expressions

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 $-3(5x - 6)$?

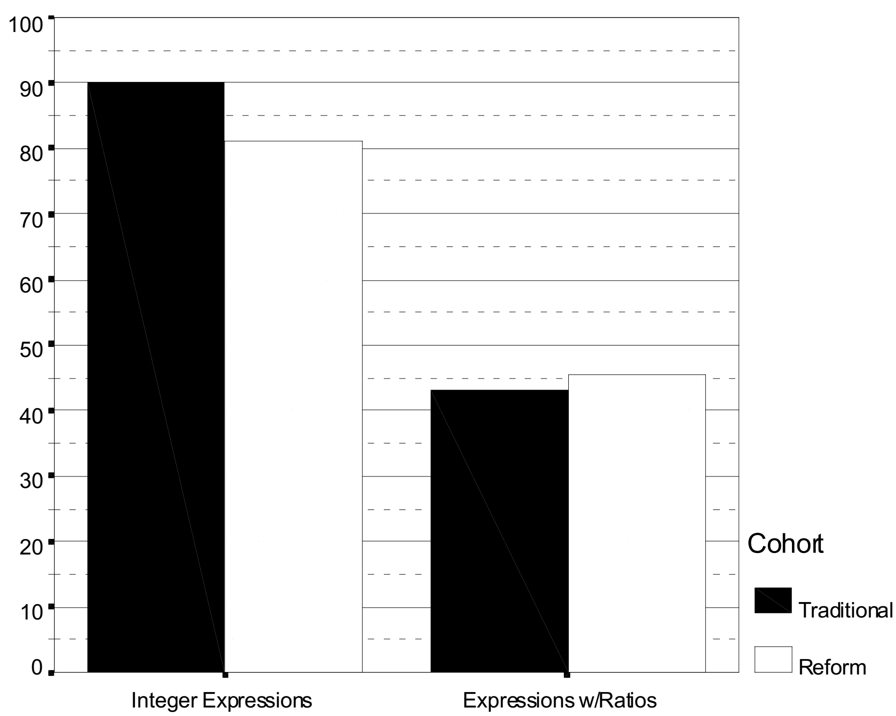
Expressions Involving Ratios

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



Raw (unadjusted for prior ability) differences between students in the Reform cohorts and students in the Traditional cohort on the remaining four sub-skills listed in   are displayed in Figure 11, Figure 12, Figure 13, and Figure 14. Differences between groups on these skills were not statistically significant. Nonetheless, all four figures show a trend for students in the Reform cohorts to score higher than students do 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

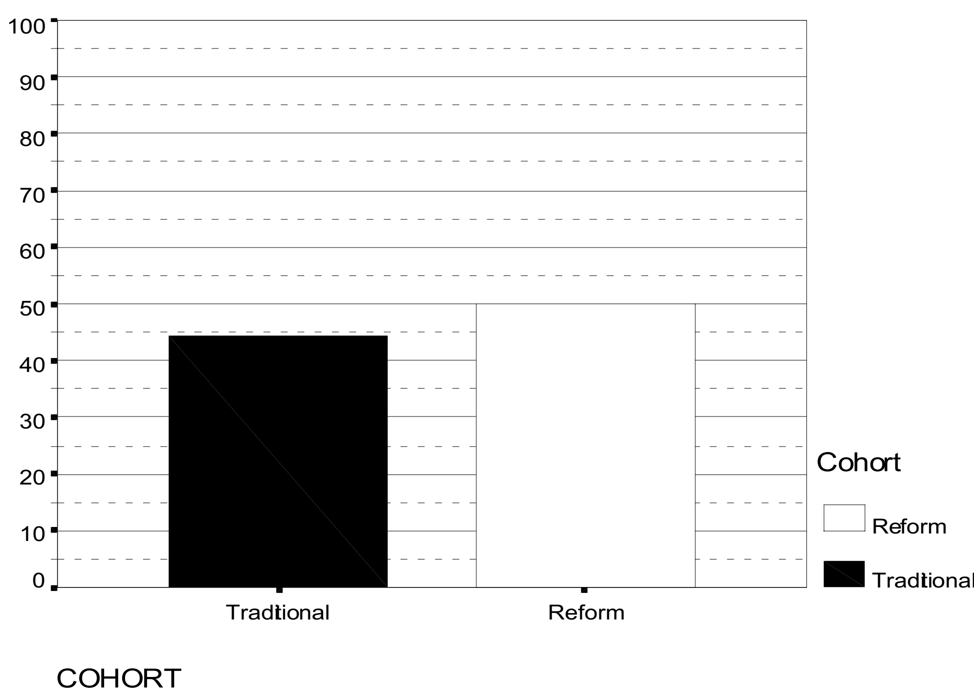


Figure 12. Mean Score Evaluating Expressions

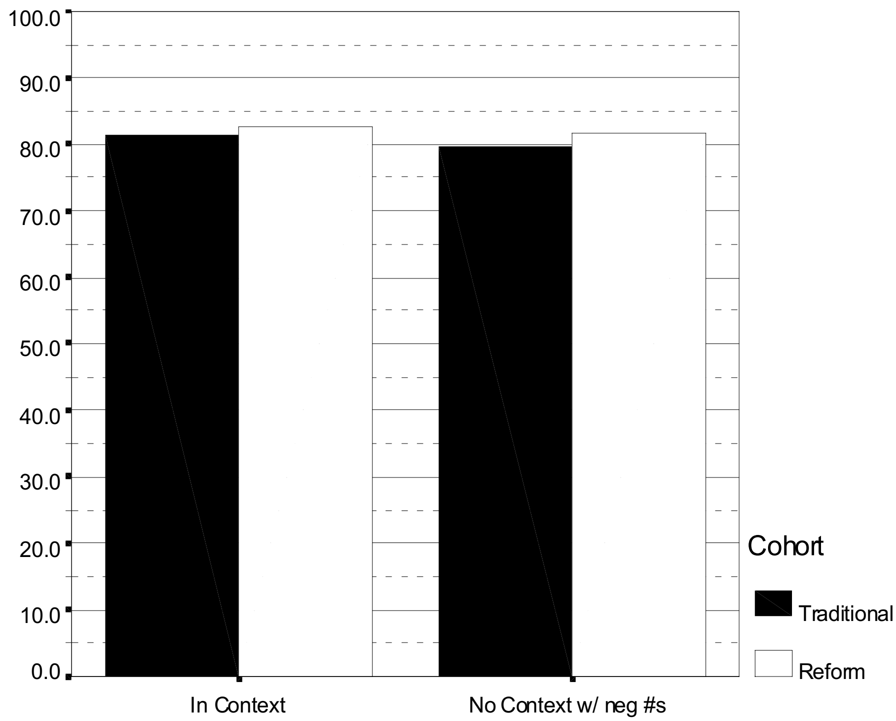


Figure 13. Mean Score Solving Linear Equations

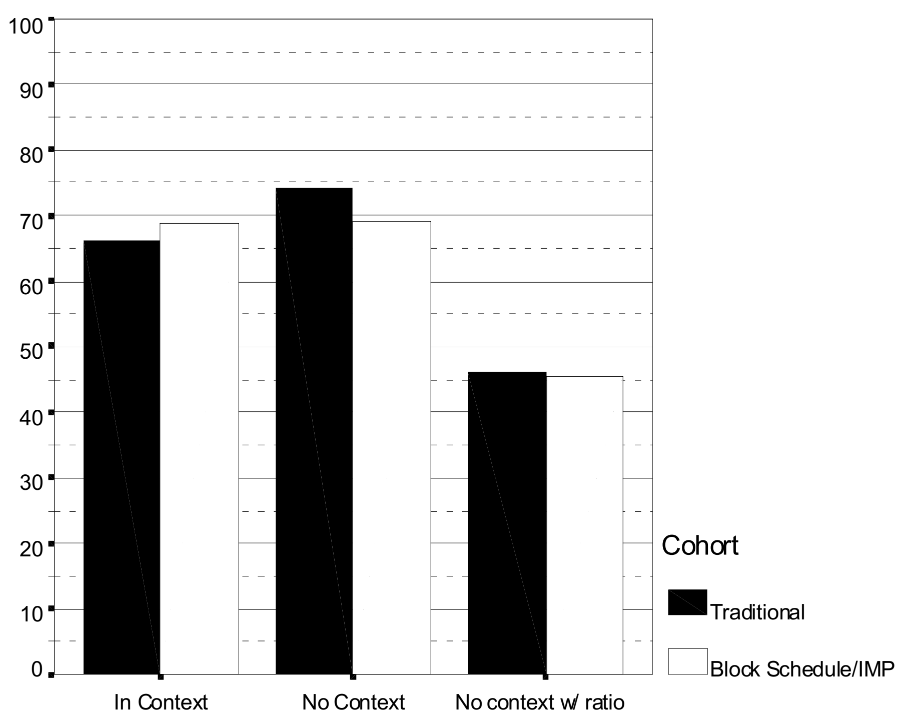
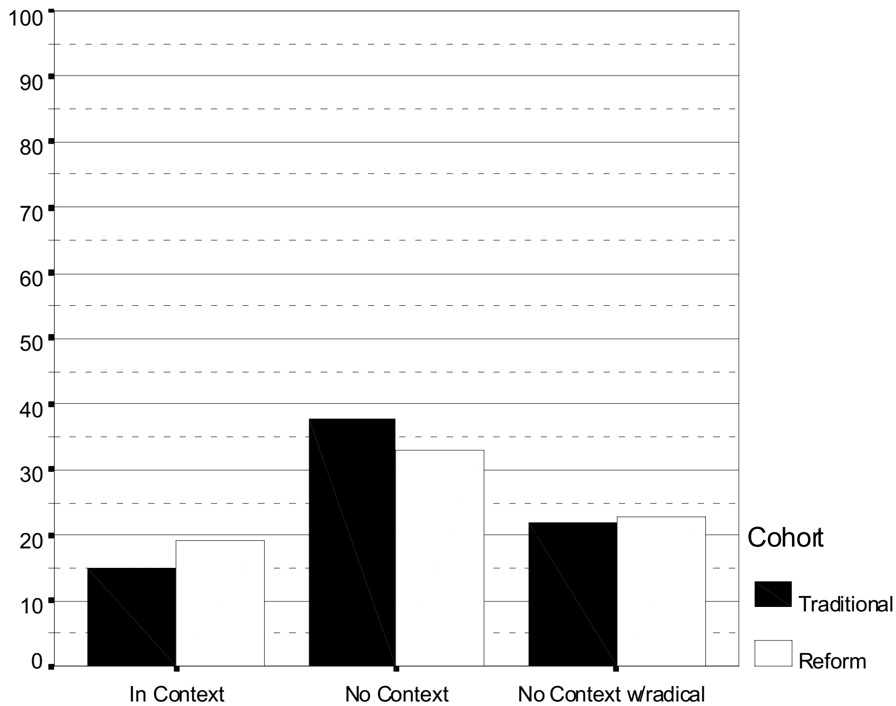


Figure 14. Mean Score Solving Quadratic Equations



<u>Results from Algebra Achievement Test 3.</u> Test 3 required students to apply multiple algebraic skills, working in pairs to solve the extended, applied algebra problem shown in Figure 15.

Figure 15. Test 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.

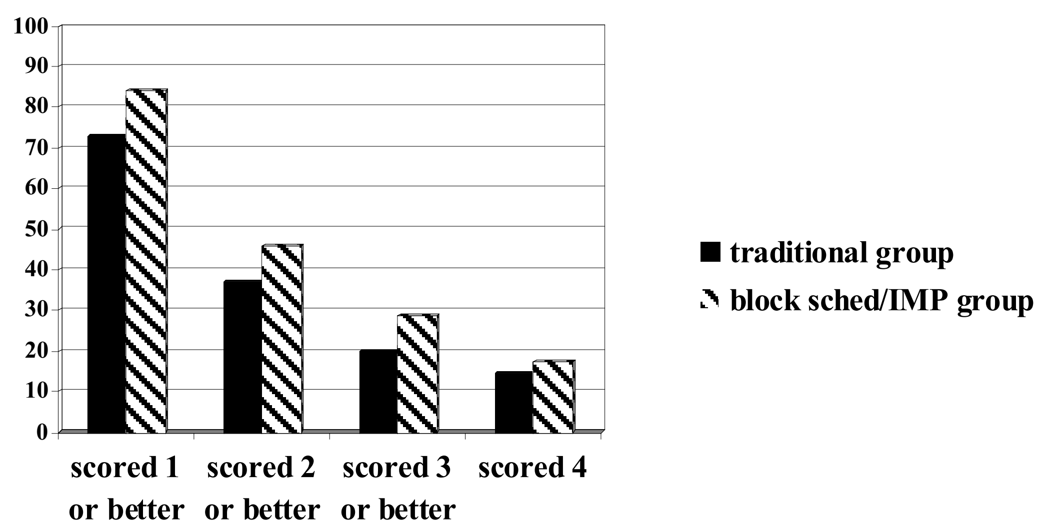
Question: For the *CD Club* the problem is to choose the long-distance service that is least expensive for their business. What advice would you give about the phone company to choose and why?

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

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 whose answer indicated they had correctly completed all four of these subtasks were assigned a score of "4"; pairs who completed three subtasks correctly received a "3", pairs who completed two subtasks correctly received a "2", and pairs who completed at least one subtask correctly received a "1". 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 .36 standard deviations in a pair's score, p=.027, which approaches statistical significance. (As noted previously, a Bonferonni 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 .027 confidence level equates to an overall p-value of 3*.027=.081 for the 3-part Algebra Achievement test.)

Grade 12 NAEP Blocks: Methodology

The three NAEP blocks completed by twelfth graders at Suburban High School were obtained from the National Center for Educational Statistics website in the spring of 1998. Unfortunately, the actual NAEP forms 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, as 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, the Traditional and IMP/Block groups were compared item-by-item, rather than on overall subscales. This resulted in 28 statistical comparisons. (There were 29 items across the three tests, but item 4 from block 1996-M10 was dropped 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.)

There were some students who stopped one or more NAEP blocks before completing that section of the test. Students who wrote something below a problem indicating they had reached the item but chose not to address it mathematically were coded as “off task” and the problem was marked wrong. The first problem a student left blank was also marked wrong. However, if a student left a problem blank, had been coded as “off task” or “blank” for the previous problem, and was coded as “blank” in all subsequent problems within the NAEP block, the problem was coded as “not reached” and the student’s score was coded as “missing”. Leaving such items missing ensured that only scores from students who actually had a chance to address each item were utilized in comparing 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, for each NAEP block we used a chi-square cross tabulation to analyze whether there was any difference between groups in the likelihood of a student giving up before finishing.

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. To keep the Type I error rate on the analysis of NAEP items at 5%, this necessitated using a very strict confidence level of .05/28=.0016 in order to state that the difference in scores between students in the Traditional cohort and students in the Block schedule/IMP cohorts were statistically significant.

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 not all students enrolled in the high school 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. Grade 6 scores were available for 93 of the 138 Suburban High School seniors enrolled in mathematics class in the spring of 1998, with a mean score 1.18 standard deviations above the national mean. Grade 6 scores were available for 97 of the 133 Suburban High School seniors enrolled in mathematics class in the spring of 2001, with a mean score 0.86 standard deviations above the national mean. Grade 6 scores were available for 122 of the 169 Suburban High School seniors enrolled in mathematics class in the spring of 2001, with a mean score 1.18 standard deviations above the national mean. These scores—all of which reflect the very high-achieving population attending this particular school—show that while students in the Traditional group may have had somewhat of an advantage in initial ability, overall the Traditional and Reform groups tested were roughly comparable.

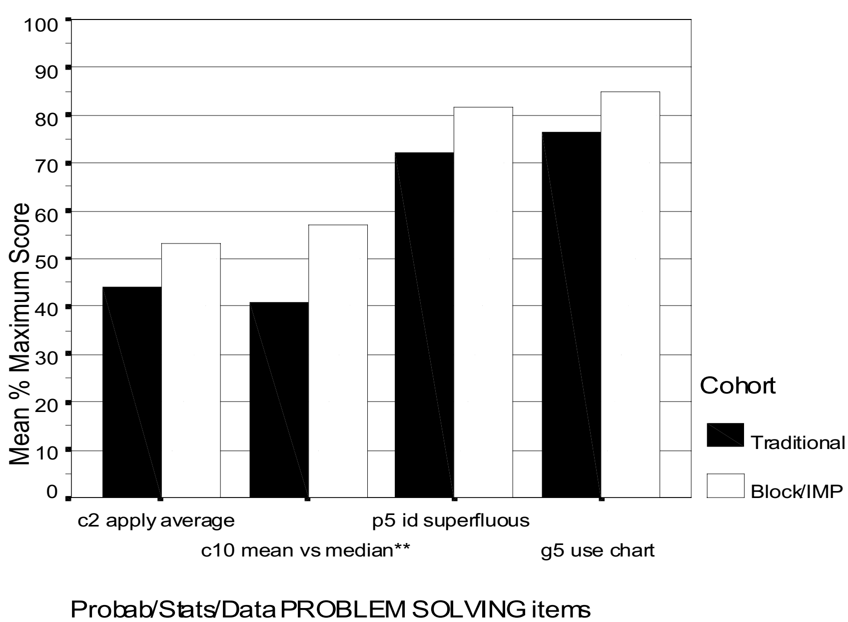
Grade 12 NAEP Block: Results

As will be discussed in more detail below, students who utilized IMP as their core curriculum under a semestered block schedule scored higher on 27 of the 28 NAEP items tested than did students who utilized a traditional core curriculum under a traditional schedule. Although it was not possible to link student scores across the three blocks to get a single "effect size," the effect size within each block was: 0.48 standard deviations for block 1996-M10 ($p<.001$); 0.46 standard deviations for block 1996-M12 ($p<.001$); and 0.26 standard deviations for block 1996-M13 ($p<.065$). Averaging these three values yields an effect size of approximately 0.40 standard deviations in favor of the Block Schedule/IMP cohorts.

NAEP items from the three blocks used for this study are categorized 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." The sections below report study results according to these classifications. For purposes of this study, the content classification "Geometry and spatial sense" was combined with "Measurement" to create one "Geometry and Measurement" content area. Similarly, the classification "Algebra and functions" was combined with "Number sense, properties, and operations" to create one "Algebra and Number" content area.

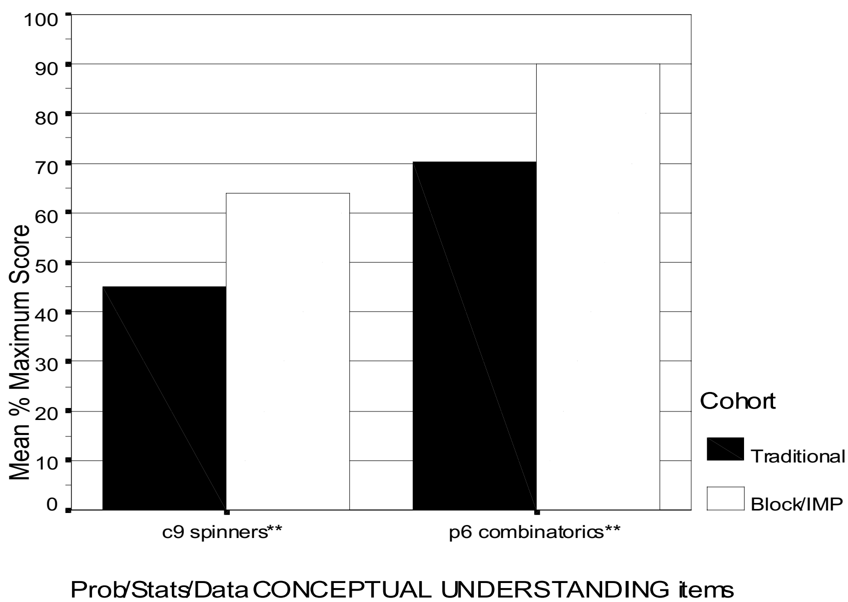
Achievement in data analysis, statistics, and probability. Across the three NAEP blocks seven items tested student achievement in the area of "Data Analysis, Statistics, and Probability." Results for problem solving items in this content area are displayed in *Figure 17.* Results for conceptual items are displayed *in Figure 18* . Results for the one procedural item dealing with Probability, Statistics, and Data Analysis are displayed in *Figure 19*.

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



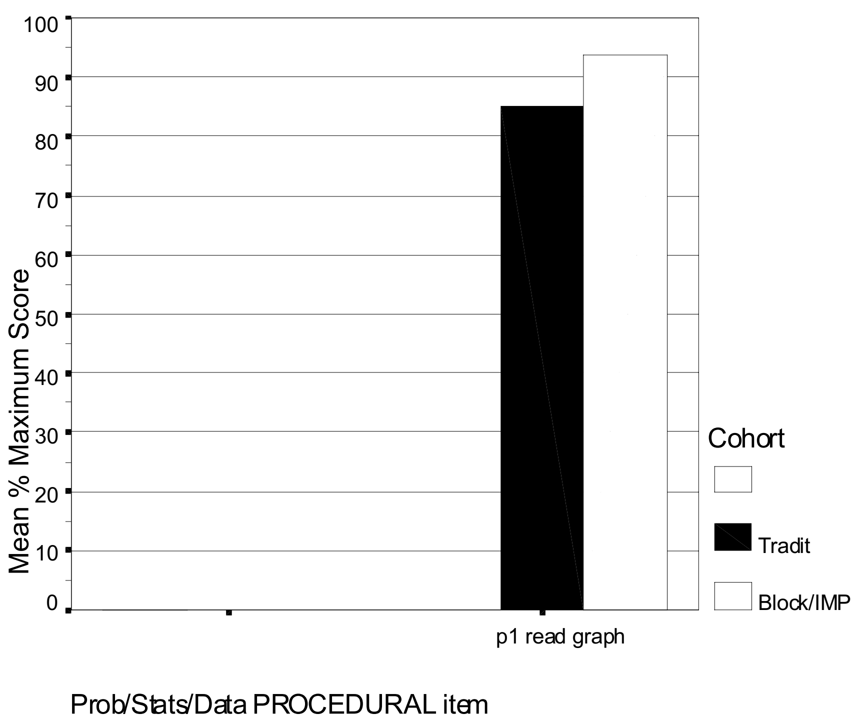
** item is statistically significant (p<.0016)

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



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

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

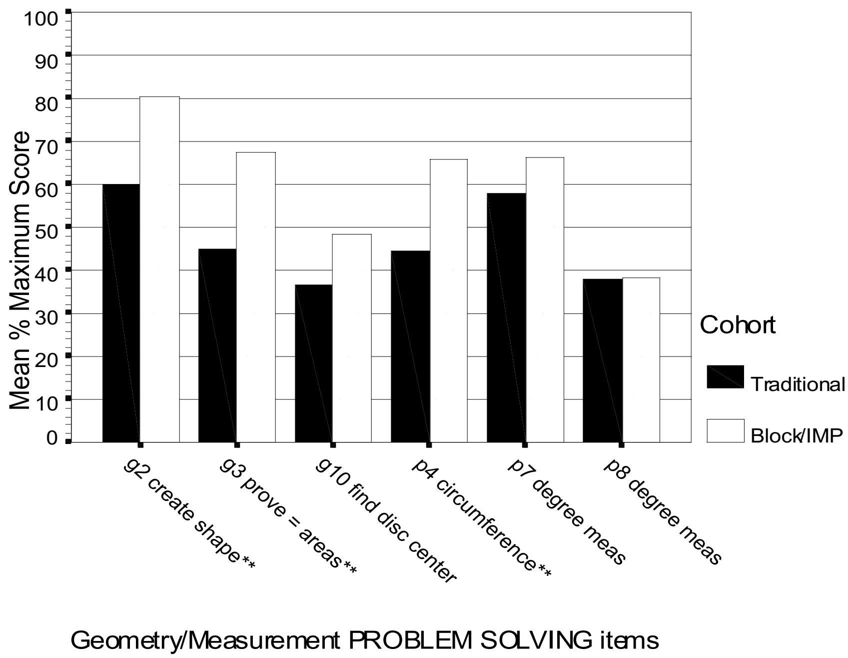


In earlier versions of the NCES website, including the one used to print out test forms for this study, the NAEP blocks used different names from what is currently the case. Block 1996-12M10 was referred to as "G3." Block 1996-12M12 was referred to as "C4." Block 1996-12M13 was referred to as "P3." These earlier designations provide a shorthand reference used in the table. For example, in the item labeled "c2 apply average" is the second item in block C (i.e., 1996-12M12), which requires students to apply the concept of average. Similarly, the item labeled "p5 id superfluous" is the fifth item in block P (i.e., 1996-12M13), which requires students to identify superfluous information.

As can be seen from the three figures, students in the Block Schedule/IMP cohorts scored higher than students in the Traditional cohort on every Data Analysis, Statistics, and Probability item tested. Further, on one of the four problem solving items and on both of the conceptual items, differences between the groups were sufficiently great to meet the strenuous .0016 confidence level required for statistical significance in this part of the study.

Achievement in Geometry and Measurement. Across the three NAEP blocks eleven items tested "Geometry and Spatial Sense" or "Measurement". *Figure 20* displays results for problem solving , *Figure 21* displays results for conceptual understanding, and *Figure 22* displays results for procedural items in these content areas.

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

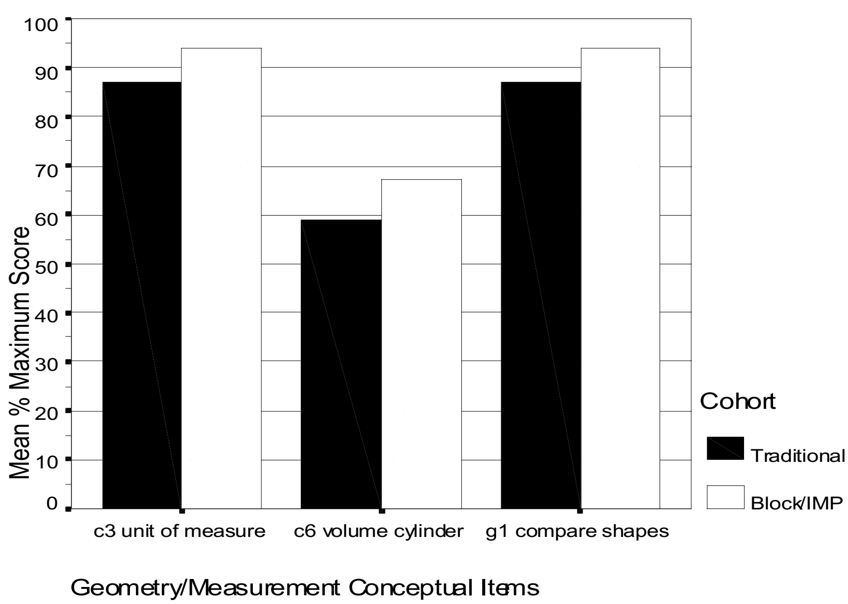
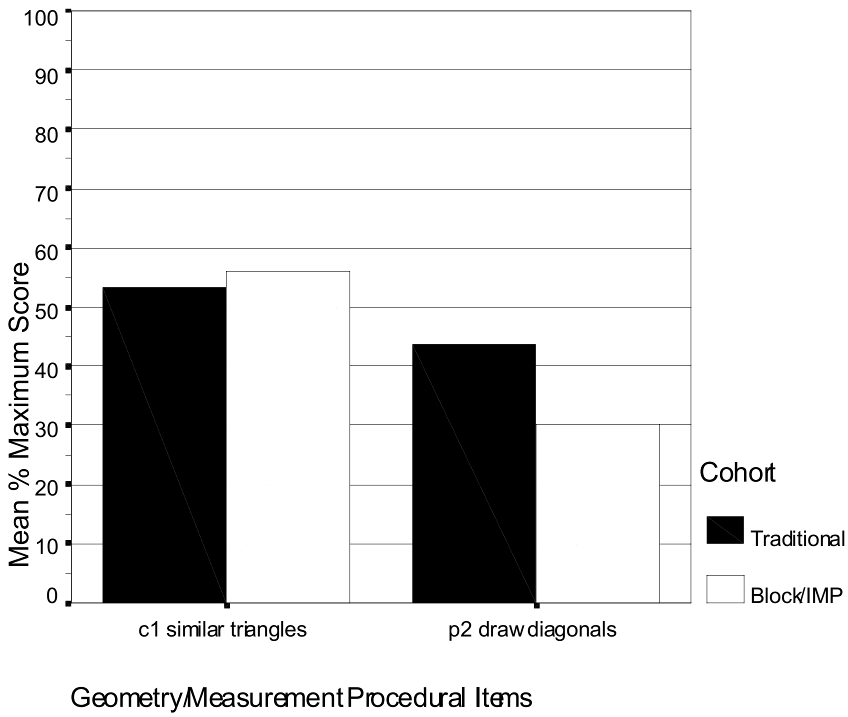


Figure 22 Procedural Items in Geometry and Measurement Content Areas



As can be seen from the three figures, students in the Block Schedule/IMP cohorts scored higher than students in the Traditional cohort on ten of the eleven "Geometry and Spatial Sense" or "Measurement" items tested, including all nine items identified by NAEP as requiring either problem solving or conceptual understanding. Further, on three of the six problem solving items the advantage of the Block Schedule/IMP group was sufficiently great to meet the strenuous .0016 confidence level required for statistical significance in this part of the study.

<u>*Achievement in Algebra and Number.*</u> Across the three NAEP blocks ten items tested "Algebra and Functions" or "Number Sense, Properties, and Operations". *Figure 23* displays results for problem solving, *Figure 24* displays results for conceptual understanding, and *Figure 25* displays results for procedural items in these content areas.

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

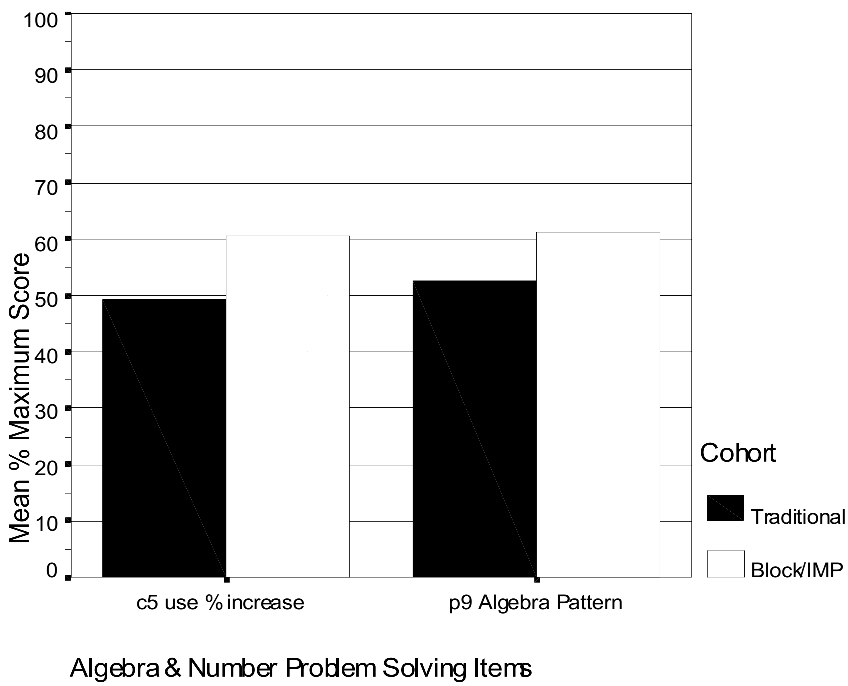


Figure 24 Conceptual Items in Algebra and Functions or Number Content Areas

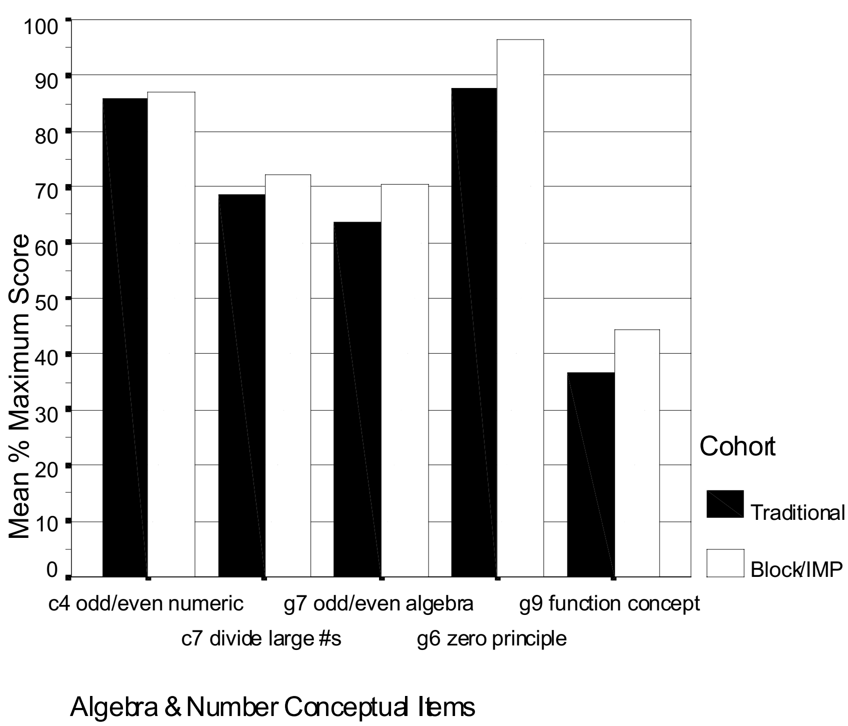
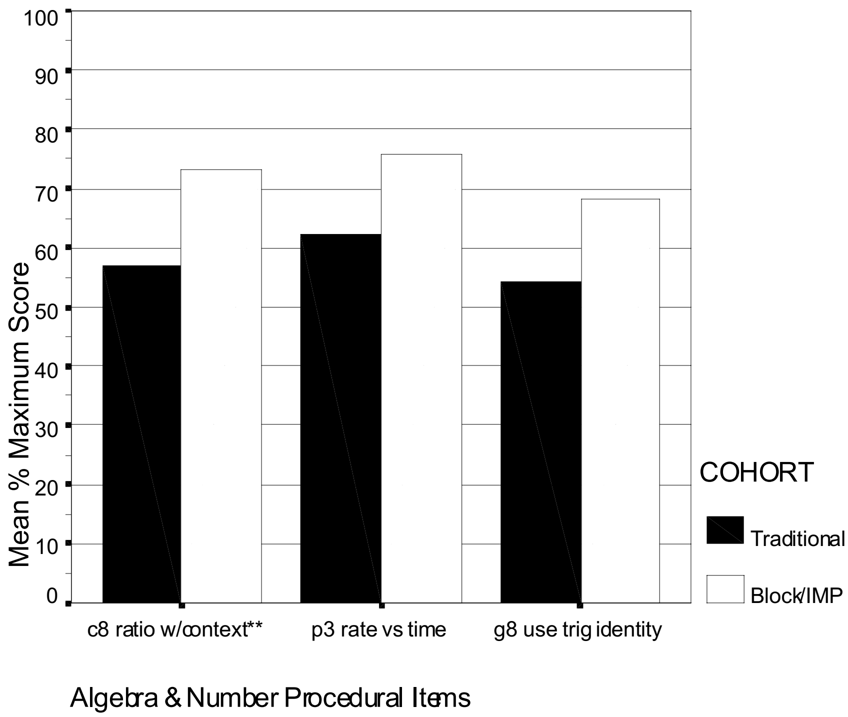


Figure 25 Procedural Items in Algebra and Functions or Number Content Areas



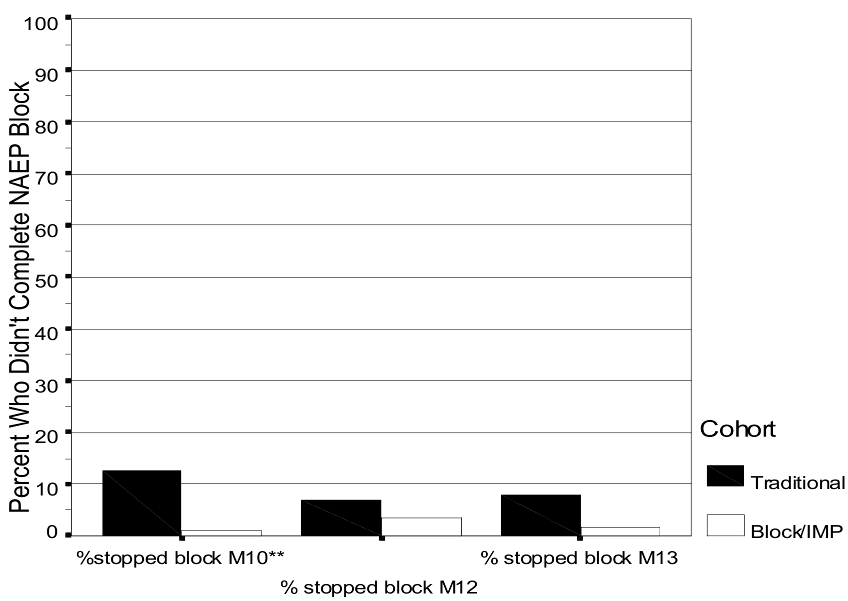
** Item statistically significant (p<.0016)

As can be seen from the three figures, students in the Block Schedule/IMP cohorts scored higher than students in the Traditional cohort on all ten items in the content areas of "Number Sense, Properties and Operations" or "Algebra and Functions". Further, on one of the three procedural items the advantage of the Block Schedule/IMP group was sufficiently great to meet the strenuous .0016 significance level.

<u>Likelihood of giving up before completing NAEP blocks.</u> 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.

As shown in on *Figure 26,* on each of the three NAEP blocks a larger percentage of students in the Traditional cohort than in the Reform cohorts failed to complete the test. Of 102 Traditional cohort students to whom we administered NAEP block 1996-12M10, 13 students (12.7%) failed to finish, vs. only 1% (2 of 200 students) in the reform cohorts who failed to finish. This difference was statistically significant ($p<.0001$). Of 101 Traditional cohort students to whom we administered NAEP block 1996-12M12, 7 students (6.9%) failed to finish, vs. only 3.5% (7 of 200 students) in the reform cohorts who failed to finish (p-value for difference=.182). Of 102 Traditional cohort students to whom we administered NAEP block 1996-12M13, 8 students (7.8%) failed to finish, vs. only 3.5% (3 of 203 students) in the reform cohorts who failed to finish (p-value for difference=.0049).

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



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

Comparing Honors students in the Pilot and Reform Cohorts

Methodology. In addition to testing seniors in the Traditional cohort during the spring of 1998 and Reform cohort seniors in the springs of 2001 and 2002, we administered the three NAEP blocks to seniors in a Pilot cohort during the spring of 2000. Like students in the Reform cohorts, students in the Pilot cohort utilized a semestered block schedule throughout high school. Unlike the Reform cohorts, the Pilot cohort used a mixture of two core mathematics curricula. Students who had begun Algebra 1 in middle school (referred to as "Honors" students) completed 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?"

Most students in the graduating classes of 2000, 2001, and 2002 had completed the CTP test in seventh grade. While equating problems prevented us from using seventh grade test scores in comparing Traditional and Reform cohort scores on the Algebra Achievement test, equating was not a problem for comparing the Pilot and Reform cohorts. Further, the mean score on the seventh grade "quantitative ability" and seventh grade "mathematics" scales proved to be a better predictor of student achievement on the NAEP blocks than was the sixth grade ability score we had used in our analysis of the Algebra Achievement test. For this reason, we used this seventh grade mean "mathematics achievement" score as a covariate when we compared the achievement on NAEP blocks of students in the Pilot cohort to those in the Reform cohorts.

In the Pilot cohort, 34 Honors students completed the three NAEP blocks. In the Reform cohorts, 63 Honors students completed all three NAEP blocks. In addition, 5 Honors students in the Reform 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 Bonferonni 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 understanding," and "Procedural knowledge." In testing each scale, we used a Bonferonni-adjusted confidence level of .017, thus maintaining an overall .017*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 also performed an OLS 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, 2002a) to perform the multiple imputation, and we utilized

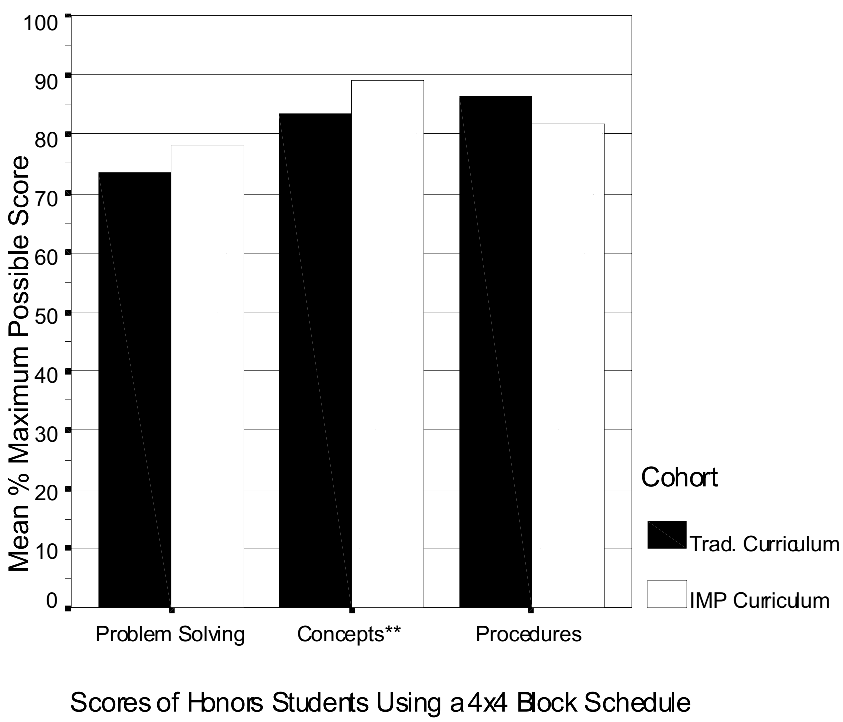
SAS Proc Mianalyze (SAS Institute, 2002b) to adjust confidence intervals to account for the greater uncertainty caused by multiple imputation of some missing data.

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 Kramer (1997) theorized might improve achievement under a semestered block schedule. They adopted a curriculum that emphasized student in-depth investigations while de-emphasizing lecture; and 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 a larger number of courses during high school. On average, Honors students in the Pilot cohort 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 Reform cohorts who took the NAEP test completed 5.6 semesters of mathematics over four years of high school. The difference in number of courses completed may be partly responsible for the achievement results reported in the next section.

Results: Changes in achievement. Figure 27 contrasts performance on NAEP items of Honors students in the Pilot cohort, who utilized a traditional core curriculum under Suburban High School's semestered block schedule with Honors students in the Reform cohorts, who utilized IMP as their core curriculum under a semestered block schedule. As the figure shows, compared to Honors students using a traditional core curriculum, Honors students using the IMP curriculum scored higher on Problem Solving and Conceptual items, and lower on Procedural items. However, of the three contrasts only the difference on Conceptual items met the Bonferonni-adjusted statistical significance requirement of .017.

Figure 27 reports differences in raw scores, unadjusted for prior ability However, it is important to confirm that the results were not attributable to differences in prior ability between Pilot cohort Honors students who completed the NAEP blocks and Reform cohort Honors students who completed the NAEP blocks. As noted above, we accomplished this by performing a probit analysis adjusting for Grade 7 CTP test scores.

Figure 27 Scores on NAEP "Mathematical Abilities" Subscales, Grade 12 Honors Students



** statistically significant (p<.0017)

After controlling for prior ability, the probit analysis confirmed that Honors students who used the IMP curriculum under a semestered block schedule scored higher on the NAEP Conceptual Understanding subscale than did Honors students who used a traditional curriculum under a semestered block schedule. The effect size was 0.57 standard deviations, and after adjusting for uncertainty due to multiple imputation the p-value was .014, which meets the

Bonferonni-adjusted criterion of "p=.017" for statistical significance. A 98.3% confidence interval for the effect size was (+0.02, +1.12).

After controlling for prior ability, the difference between the two Honors groups on the Problem Solving subscale and on the Procedural subscale remained statistically non-significant. For the Problem Solving subscale the estimated effect size was 0.31 standard deviations, p-value=0.15, 98.3% confidence interval for effect size= (-0.21, +0.83). For the Procedural subscale the estimated effect size was -0.46 standard deviations (i.e., an effect size favoring the Pilot group), p-value=0.07, 98.3% confidence interval for effect size=(-1.04, +0.14).

<u>Changes in Mathematics Course Enrollment: Traditional Cohorts vs. Reform Cohorts</u>

<u>Methodology.</u> 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. To analyze the effects of the new schedule and curriculum on student course enrollment, we decided to use as comparison "Traditional" cohorts students from the two graduating classes prior to the graduating class of 1998, the "Traditional" cohort who completed the Algebra Achievement test and the three NAEP blocks. We chose these earlier cohorts because the class of 1998 had utilized a semestered block schedule during their senior year, and would not provide an accurate picture of course-taking patterns at Suburban High School under a traditional schedule and curriculum.

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 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 20-day spring session was discontinued.

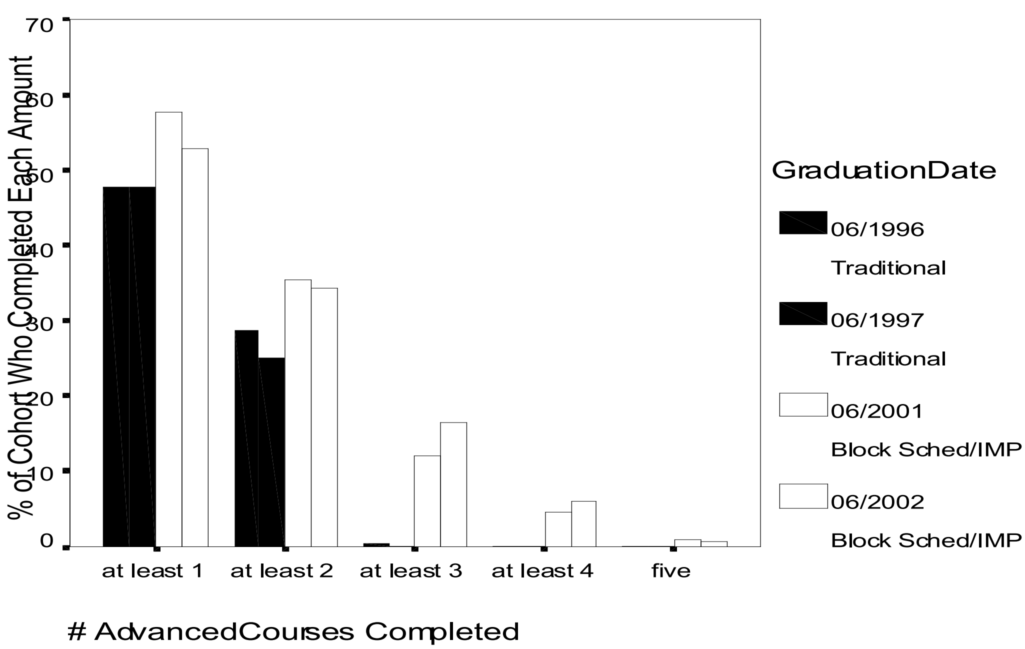
To compute the number of hours students were enrolled in mathematics classes, we multiplied the credits earned in each class times the number of allocated hours for that class, and summed this value over all mathematics classes the student took. All mathematics classes were worth one credit, except for 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 _ mathematics credit.

We also report the number of "Advanced" courses students completed. In general, 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. There were two exceptions to this rule. We did not count "SAT Prep Math" as an advanced course. Also, we did not count a course named "Algebra 3/Trig" as an advanced course. Algebra 3/Trig was a course taken by 64 students in the graduating class of 1996 and 65 students in the graduating class of 1997. The course was designed for students who had received low grades in Algebra 2 or who had completed a less advanced 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). *Advanced Algebra* is the fourth of six mathematics texts in a Grade 7-12 series designed by the University of Chicago School Mathematics Project (UCSMP), and is recommended for all high school students whether or not they are college-intending. In general, the topics covered in *Advanced Algebra* are less advanced than those in the fourth IMP course in either the College Preparatory or in the Honors track. Because we did not count Integrated Math 4 College Preparatory (completed by 106 students in the First Reform cohort and 138 students in the Second Reform cohort) or Integrated Math 4 Honors (completed by 84 students in the First Reform cohort and 73 students in the Second Reform cohort) as "advanced", we felt that we could not count Algebra 3/Trig as "advanced." By these criteria, after completing Algebra 2 students in the two Traditional cohorts could enroll 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 advanced courses that had been offered to the Traditional group except for the Honors version of Algebraic Analysis, which was dropped from the curriculum. Also, under the semestered block schedule, four new advanced mathematics courses were added to the Suburban High School curriculum. The new courses were: Discrete Analysis, Discrete Analysis Honors, Statistics, and Statistics AP. Finally, 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/curriculum graduating class of 1996 enrolled in 501.8 hours of mathematics class and the average students in the traditional-schedule/curriculum graduating class of 1997 enrolled in 494.2 hours of mathematics class. 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 in mathematics class. 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 class.

The increased enrollment was particularly evident in advanced classes such as statistics, discrete analysis, precalculus and calculus. Figure 28 shows the actual distribution of students enrolled in various hours of advanced mathematics courses. As the figure shows, the percentage

Figure 28. Number of Advanced Course Completed.



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.

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 BC Calculus exam grades from these School Profiles.

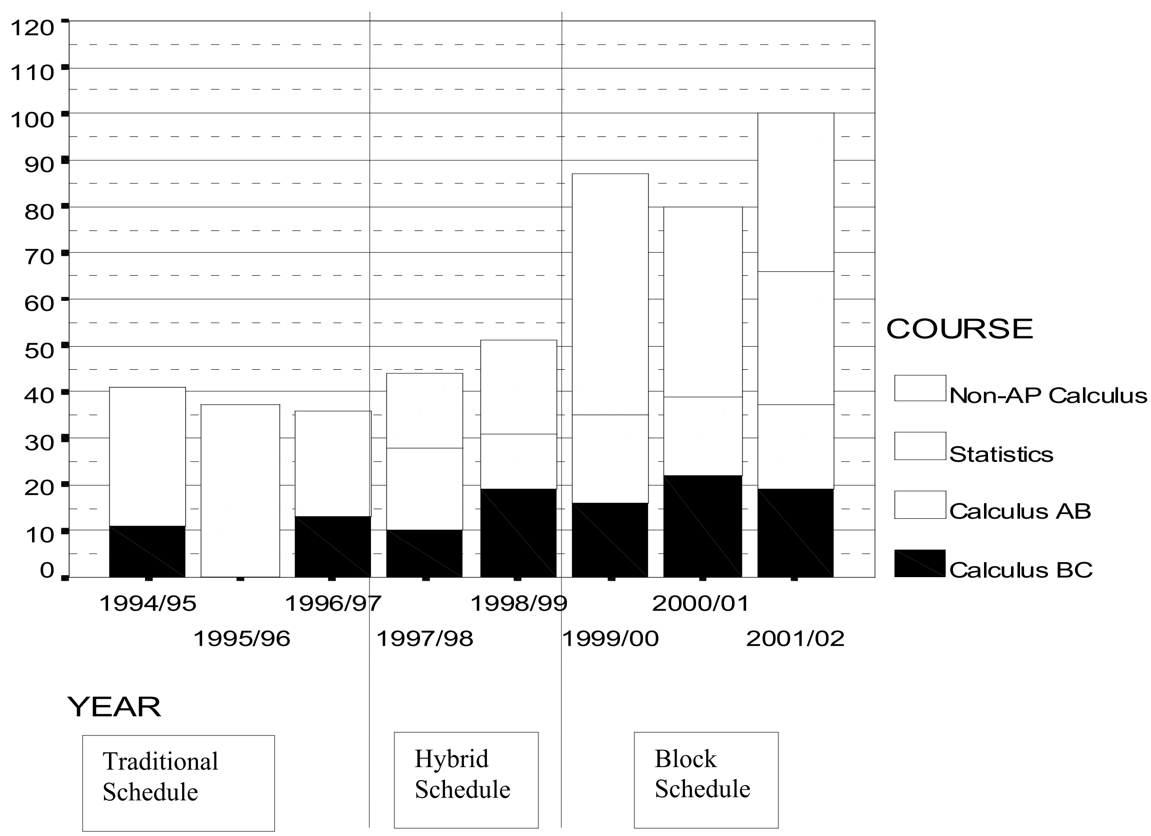
Results. Adoption of the semestered block schedule had a direct impact on the way Advanced Placement (AP) mathematics classes were taught. Previously, Suburban High School had offered students a choice of two AP mathematics classes: either a yearlong Calculus AB course, or a more rigorous yearlong Calculus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

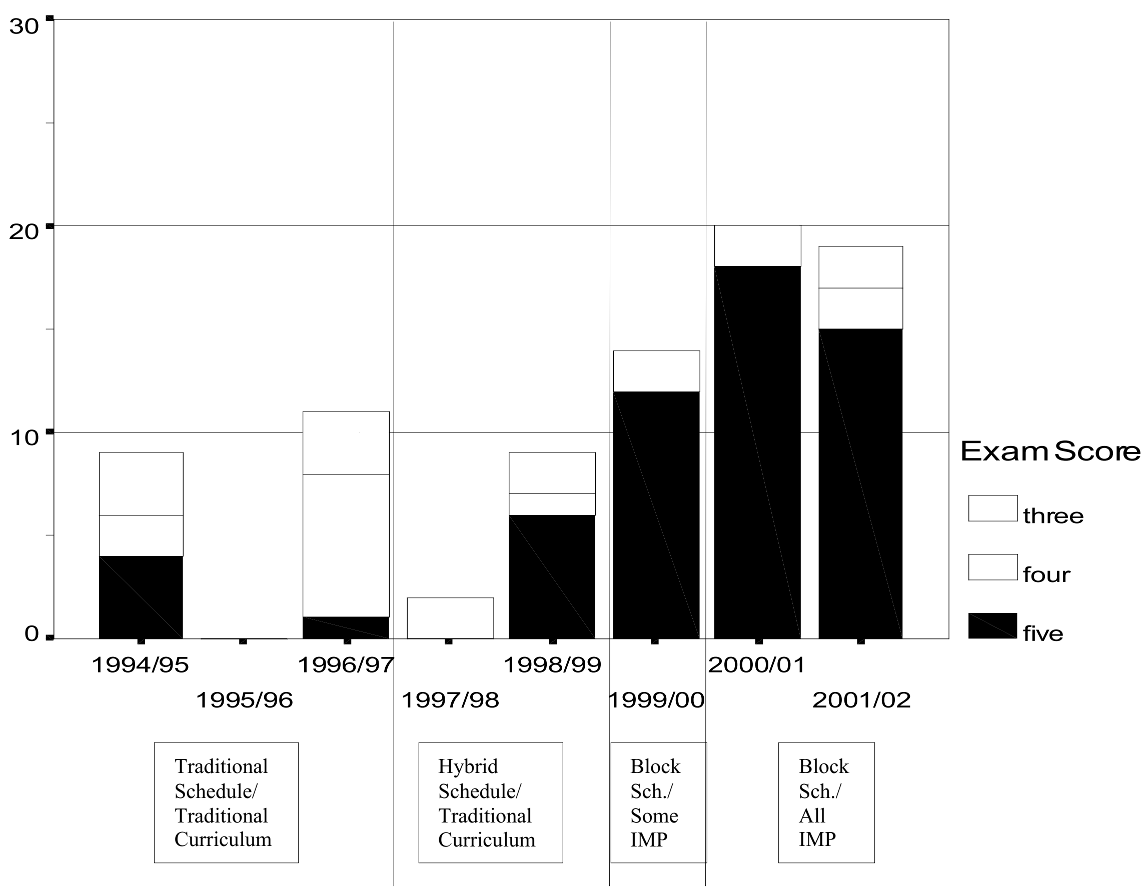
Figure 29 displays the way course enrollment changed as the semestered block schedule was implemented. Note that in the "hybrid schedule" years of 1997/98 and 1998/99 Advanced Placement students had utilized a traditional schedule for their early high school years but switched to a semestered block schedule 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.



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 a senior 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 theorized 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, at translating an exponential function from tabular form to an equation. They also appeared to be better 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 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 availed themselves of 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 Reform cohorts were achieving better than students in the Traditional cohort in all areas of mathematics tested. These included the content areas of "Data Analysis, Statistics, and Probability," " Geometry and Measurement," and "Algebra and Number." Within each content area, students in the Reform cohorts achieved higher than those in the Traditional cohort on Problem Solving, Conceptual Understanding, and Procedural items.

While achievement results measured by 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 achievement results measured by 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 Reform 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 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 13.3% 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.

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 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.” Another example is Item 8 from NAEP Block 1996-12M12, an algebra/number procedural item on which the advantage of Reform cohort students over Traditional Cohort students was statistically significant. That item read, “Luis mixed 6 ounces of cherry syrup with 53 ounces of water to make a cherry-flavored drink, Martin mixed 5 ounces of the same cherry syrup with 42 ounces of water. Who made the drink with the stronger cherry flavor? Give mathematical evidence to justify your answer."

Additional data on the interaction between schedule and curriculum was provided by NAEP tests completed by Honors students in the spring of 1998. These tests enabled us to compare the achievement of high-ability students from that year, who utilized a traditional sequence of Algebra 1 Honors in eighth grade followed by Geometry Honors and Algebra 2 Honors under a semestered block schedule, with Honors students from later cohorts, who utilized an IMP sequence as their core curriculum in middle school and under the high school's semestered block schedule. Results indicated that Honors students who had utilized IMP were superior in Conceptual Understanding, with an effect size of +0.56 standard deviations on Conceptual Understanding items from the three NAEP blocks. While the IMP students also scored higher on Problem Solving items and the traditional curriculum students scored higher on Procedural items, neither of these latter two contrasts approached statistical significance.

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 it had been under the traditional curriculum and schedule.

This current study was not a controlled experiment of 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 a larger number of courses, teachers at Suburban High School decided to have their top students complete the four IMP texts, followed by a one semester precalculus 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, combining a semestered block schedule with the problem-centered IMP curriculum appears to have succeeded in improving student mathematics achievement at Suburban High School. It would be worthwhile to pursue this promising combination of reforms in other settings, to see if the positive results can be replicated.