Chapter 2: Review of the Literature

The idea for conducting an investigation into the joint effects on mathematics achievement of a semestered block scheduling and an NCTM *Standards*-based mathematics curriculum began with a review of the literature on block scheduling that was published in the *Mathematics Teacher* in December, 1996 (Kramer, 1996). Based on both the literature reviewed and on interviews with teachers, that article noted that lecturing appears to work less well under a block schedule than under a traditional schedule. As a consequence, researchers generally recommended including a larger percentage of participatory activities in each block-scheduled class period. The article concluded with the following observation:

> If block scheduling were implemented with adequate planning time and staff development, and with administrative policies that maintained the number of classroom hours allocated to mathematics over a student's high school career, it is quite possible that achievement would be higher than it had been under a traditional schedule. To date, unfortunately, such an implementation has not been studied.

In response to the article in *Mathematics Teacher*, Mrs. Sullivan[2], the mathematics supervisor at Suburban High School contacted the current author. She indicated that her school was in process of adopting a semestered block schedule and was implementing all of the recommendations contained in the *Mathematics Teacher* article. Further, they were simultaneously adopting IMP, a problem-centered curriculum designed to implement the NCTM *Curriculum and Evaluation Standards* (NCTM, 1989). She believed that IMP was particularly well suited to a block schedule and was likely to further enhance student achievement. She invited the author to help her investigate the

[2] "Mrs. Sullivan" is a pseudonym

achievement effects of the new schedule and curriculum at Suburban High School. Thus the current study was born.

This Chapter summarizes and updates the original block scheduling literature review published in 1996. It also reviews the literature on IMP and on the achievement effects of other high school curricula designed to implement the NCTM *Standards*.

Semestered (4x4) Block Scheduling

Block scheduling is not a new phenomenon. It has been widely used in British Columbia, Ontario, and Alberta since the 1970s. In the United States block schedules became increasingly popular throughout the 1990s, and today have spread to high schools in all regions of the nation.

This study focused on one of the two most common types of block schedule, namely, the semestered or 4x4 block schedule. A semestered block schedule typically consists of four courses meeting 80-90 minutes daily for 90 days. The other common type of block schedule is the alternating-day**,** or A/B block schedule. An alternating day block schedule typically consists of 8 courses meeting 80-90 minutes every other day for the entire 180-day school year. Besides these two most common types of block schedules, other forms of block scheduling have been tried in various places. The most important are “interdisciplinary” block schedules (Sigurdson, 1982) and “quarter” plans (a more intensive extension of semestered schedules, in which students usually take two academic courses at a time, with each course meeting about 150-180 minutes daily for one quarter of the school year.) Canady and Rettig (1995) provide an extensive discussion of various forms of block scheduling.

Administrators are often attracted to block schedules for a number of reasons. There is evidence that student discipline improves under all major forms of block scheduling (see, e.g., Carroll, 1994; Hackman, 1995; Hillcrest High School, 1995; Meadows, 1995; Reid, 1995a; and Sessoms, 1995), as do student attitudes towards school (Averett, 1994, Kramer 1997a, 1997b, Meadows, 1995, Stevens, 1976). Semestered and other intensive forms of block scheduling (e.g., the “quarter” plan) appear to lead to reduced dropout rates (Kramer, 1997a; Sharman, 1990). Finally, administrators often gain flexibility in scheduling by having each student take eight courses per year under a block schedule, versus six or seven courses per year under a traditional schedule.

Effects of Semestered Block Scheduling on Mathematics Instruction

Reduced effectiveness of lecturing. In general, the literature reviewed indicated that teaching by lecture alone works less well in a longer time block (Howard High School, 1994; King, Clements, Enns, Lockerbie, & Warren, 1975; King, Warren, Moore, Bryans, & Pirie, 1978; Meadows, 1995; O'Neil, 1995; Reid, 1995a; Sturgis, 1995). Not surprisingly, students seem to find it difficult to sit through a class that consists essentially of two lectures conducted in sequence. Instead, researchers generally recommended including a larger percentage of participatory activities in each block-scheduled class period.

The literature available, however, has two weaknesses. First, the conclusion has not yet been confirmed by student performance data. All of the literature that recommended placing additional emphasis on non-lecture teaching modes based its conclusions on surveys of administrators, teachers, and/or students.

Second, the literature indicates that lecturing is less effective under a block schedule for all subjects in general; it does not address mathematics in particular. In fact, there is some evidence that mathematics teachers may be less likely to change their teaching methods under a block schedule than are teachers in other departments. Reid (1995a) interviewed five principals of schools in British Columbia that had switched to an intensive (four quarters, two courses per quarter) block schedule. He found that mathematics teachers in these schools had a harder time changing their teaching methods than did teachers from other departments.

The need to adopt new teaching modes is reflected in a comment that was made repeatedly by mathematics teachers interviewed by Kramer (1996): Experienced teachers said they "felt like first year teachers" after switching to a block schedule. Apparently, pedagogical methods that teachers had learned from experience in traditional classrooms did not translate successfully into block scheduled classrooms.

Thus, it is reasonable to draw the tentative conclusion, based on the opinions of teachers, administrators, and students, that teachers in mathematics need to reduce their amount of lecturing in order to maintain student interest under a block schedule. Furthermore, mathematics teachers are likely to find such a change more difficult than are teachers in other subject areas, and thus may have a particularly strong need for staff development and extra planning time to assist them in making the change (Brophy, 1978;

King, et al., 1978; Watts & Castle, 1992; Averett, 1994; Meadows, 1995; Reid, 1995a; Canady & Rettig, 1995; Salvaterra & Adams, 1995; Kramer, 1997a).

Decreased breadth and increased depth of coverage. A literature review by King et al. (1975) reported that both mathematics and French teachers experienced particular difficulty covering the equivalent of two classes of material during a double length period. A follow-up study made detailed observations of classrooms in six schools with semestered block schedules. Comparing mathematics classrooms in these schools to ones they had observed operating under a traditional schedule, they noted that under a block schedule mathematics teachers frequently used up more instructional time to cover the same content (King, Warren, Bryans, & Pirie, 1978).

More recently, Usiskin (1995) reported the opinions of several teachers who had taught under both traditional and alternating-day block schedules using textbooks developed by the University of Chicago School Mathematics Project. All agreed that under the alternating-day block schedule, not as much content was covered. However, it is possible that Usiskin's results can be explained by the fact that there are fewer allocated hours per course. Sturgis (1995) reported that some teachers at a school in Maine were covering less content after switching to an alternating-day block schedule, but that this was balanced by an opportunity to go into more depth. While the data collected by both Usiskin and Sturgis related to alternating day block schedules, it is likely that their observations apply to semestered block schedules, which share the characteristics of more hours per class period combined with fewer allocated hours per course typical of an alternating day schedule.

A number of surveys support the observation that teachers perceive the longer time blocks as providing an opportunity to teach concepts in greater depth. This is true for both semestered and alternating-day block schedules. A survey of teachers at 21 North Carolina schools that were in their first or second year of implementing a semestered block schedule reported that over 70% of those surveyed felt that the block schedule had a "moderate positive effect" or "strong positive effect" on each of the following: 1) problem solving ability, 2) higher order thinking, 3) in-depth knowledge, and 4) retention of subject matter (Averett, 1994). A survey of teachers at four Maryland high schools in their first or second year of implementing a semestered block schedule obtained similar results (Meadows, 1995). Other studies have reported increased depth of coverage by teachers at schools using an alternating-day block schedule (Sessoms, 1995; Sturgis, 1995).

Need to adjust the mathematics curriculum. A number of researchers have noted that faculties need to adjust the curriculum under a block schedule (King, et al., 1975; Harter, 1994; Sessoms, 1995). Administrators often move to a block schedule in order to enable students to take a larger number of courses each year, while allocating fewer hours per course. For example, a school that had previously offered students seven 45-minute periods daily might move to a schedule that offered students four 80-minute periods daily. Under a semestered block schedule, each course would then meet for half the year. Students would be replacing two 45-minute periods with one 80-minute period, but would be taking one extra course yearly.

Under this type of schedule, students often enroll in a larger number of core courses (Cameron, 1995), and in particular in a larger number of mathematics classes (Edwards, 1995). One study reported that students enrolling in additional courses were often of two types: students who had failed a class and were retaking it, and top students who were taking two mathematics courses a year (Kramer, 1996).

Harter (1994) noted this kind of schedule allows schools to offer students more time to take mathematics, not less. He emphasized that students unlikely to succeed within existing time constraints could benefit from two-term core mathematics courses, that honors programs could offer two semesters of challenging mathematics yearly to interested students, and that all students could benefit from courses in statistics and data analysis.

However, administrative constraints can make it impossible for schools to take advantage of this opportunity. Harter (1994) reported that principals in North Carolina's block scheduled schools were often allotted faculty slots within distinct classifications (i.e., so many mathematics teachers, so many science teachers, so many art teachers, etc.) and that these allotments could result in an imbalance between student needs and faculty available to teach particular courses. According to Harter, this resulted in many schools on the four-period day (i.e., with semestered block schedules) encouraging students to take elective courses outside the core disciplines, while discouraging or even denying some students access to two courses per year in the same academic discipline. He concluded that the potential advantages of a block schedule could be realized fully only if high school principals were given more flexibility in assigning teaching positions. Kramer (1996) interviewed eight teachers from schools having success with block scheduling. Seven reported adjusting the mathematics curriculum to accommodate the new schedule. Curriculum changes mentioned by teachers in the sample included the following:

- creating a 2-part algebra class for "lower" mathematics students;
- replacing the normal Algebra I/Algebra II sequence with a sequence of three (shorter) Algebra courses;
- modifying Geometry and Algebra I courses to eliminate topics taught in Algebra II;
- splitting a combined Algebra II/Trigonometry class into two separate classes;
- adding new courses (e.g., in statistics) for students who finish up the regular sequence.

A good example of the potential importance of reworking curriculum while moving to a block schedule is seen in the following description, sent in response to a letter to the editor of Mathematics Teacher:

> Our school is in its third year of using the block schedule where students complete one course in 18 weeks. The only problem we had was with Algebra. All of the other courses were easily adapted to the 85 minute class, but algebra in 18 weeks is just too fast. We met with such failure the first year that our administration readily went along with changing algebra to a full year class (36 weeks). Algebra is the foundation of all of our other classes and the students need to have a solid foundation before we can expect them to succeed in following courses. The full year classes in algebra allow plenty of time to do the exploration and hands on activities that help the students get a better understanding. In the 18 week algebra, it was practically impossible to do any activities--we felt like we were flying thru the material and losing lots of students in the process. There's not a lot of algebra that can be dropped (chopped, condensed, combined, etc.) without adversely affecting future courses.
>
> Since changing to 36 weeks of 85 minutes of algebra each day, we have had a much better success rate and the geometry and algebra II teachers have noticed a difference. (Kramer, 1996, pp 760-761).

Advanced placement classes. Advanced placement (AP) classes present special challenges to schools switching to block schedules. One teacher summarized the situation as follows:

> AP exams are given in May only. Students who take the AP course first semester are rusty, students who take the AP course second semester won't cover enough by May (Kramer, 1996, p. 761).

In some cases, schools using block schedules offer Calculus and other AP courses as double-length courses that run the entire year (Edwards, 1995; Governor Thomas Johnson High School, 1995) or three quarters of the year (Schoenstein, 1995), with the last quarter perhaps offering a special topic, like a probability and statistics class. One article described a block scheduled school that switched its AP courses back to standard 45-minute classes. For example, AP English and AP social studies ran in back-to back 45-minute classes for the entire school year. The authors noted, however, that several teachers thought this was a step backwards (Salvaterra & Adams, 1995).

The College Board (1998) reported that in 1998 four instructional schedules were widely used for AP courses:

1. traditional schedule of 30- to 60-minute sessions each day during the school year
2. schedule of 61- to 90-minute sessions every day during the school year
3. semesterized fall block course
4. semesterized spring block course

Schools using semestered block schedules have the option of offering Calculus and other AP courses in any one of the latter three formats: as double-length courses running all year, in the fall, or in the spring. The College Board compared achievement on the 1997 AP Calculus AB exam for each of these four types of schedule, using a Bonferroni-adjusted significance level of .08 for each of the six possible pair-wise comparisons. They found that, after adjusting for prior ability as measured by scores on the mathematics portion of the PSAT/NSMQT administered in 1995 or 1996, student AP scores could be predicted by how much time they had spent studying the topic. Students in double-length courses running all year scored significantly higher than students in year-long 30- to 60-minute courses, who in turn scored significantly higher than students enrolled in AP Calculus AB for just the spring semester or just the fall semester. Scores of students enrolled in spring-semester or fall-semester courses were not significantly different from each other.

Impact of student absences. A student who misses a day under a block schedule misses nearly twice as much lesson time. Thus, teachers have indicated that absences are more disruptive to student learning under both semestered and alternating-day block schedules than they are under traditional schedules. A majority of North Carolina teachers responding to Averett's (1994) survey indicated that, under a semestered block schedule, their students had difficulty in recovering from absences. This, along with difficulty in accommodating transfer students, was one of the two major weak points they noted.

Usiskin (1995) reported similar opinions among teachers using materials from the University of Chicago School Mathematics Project in an alternating-day block schedule. Further, Sturgis (1995) reported that an alternating-day block schedule made it more difficult for teachers to ensure students made up missed homework after an absence.

Retention of learning after a gap in sequential instruction. Under a semestered block schedule, students often enroll in mathematics during only one of the two yearly semesters. Parents and teachers have often expressed the concern that mathematics achievement might be harmed because of students' inability to retain information when the gap between one mathematics course and the next one could be more than one year (Lindsay, 2002).

Joseph Carroll, one of the earliest and best-known proponents of block scheduling, provided data relevant to this question. He reported the results of nine tests that compared achievement of students with a gap between learning and test taking of three months to that of students with a gap of six months; as well as nine tests that compared achievement of students with a gap between learning and test taking of three months to that of students with a gap of nine months (Carroll, 1994, p. 11). Although Carroll claimed that levels of retention were comparable between students with longer and shorter gaps, his data appear to show otherwise. Two of the nine tests that compared a three-month gap to a six-month gap showed statistically significant higher scores after the shorter gap; four of the nine tests that compared a three-month gap to a nine-month gap showed significantly higher scores after the shorter gap. Differences on the remaining thirteen tests were not statistically significant. Carroll also reported six tests that compared achievement of students with a gap between learning and test taking of six months to that of students with a gap of nine months or twelve months, and found no significant differences on any of the six tests. It should be noted that Carroll did not provide details on these tests, so it is unclear what academic subjects displayed the effects, or even what p-value was used to judge "significance."

The potential impact of a gap in sequential instruction was addressed in a 3-year longitudinal study involving 214 students in London, Ontario who completed their ninth-grade mathematics course in 1972. Of these students, 107 studied tenth-grade mathematics in all-year schools, 63 studied tenth-grade mathematics in the first semester (fall of 1972), and 44 studied tenth-grade mathematics in the second semester (spring of 1973). At the end of their ninth-grade year (i.e., June, 1972), all students were given a 28-item test, consisting of a 10-item Numerical Skills subtest and an 18-item Algebraic Skills subtest. The three groups scored nearly identically on both subtests.

Each group was administered the same test at the beginning of their tenth-grade mathematics course. Thus, the 44 second semester students had a longer gap (summer-plus-fall) before beginning instruction than did students in the other two groups (summer only). Although there were no differences among the three groups on the Basic Skills subtest, the second semester group (i.e., the group with the longer time gap) scored lower than the other two groups on the Algebraic Skills subtest.

The test was administered again at the end of tenth-grade instruction: January, 1973 for the first-semester group, and June of 1973 for the all-year group and the second-semester group. By the end of tenth-grade instruction, the second semester group had caught up with the other two, so that there were again no differences in test scores on either subtest. Finally, all three groups were administered the test in fall of 1973 (i.e., the beginning of eleventh grade), and all three maintained their scores, with the groups receiving nearly identical results on both subtests.

Thus, when tested, students with an extra semester time gap did have more difficulty recalling recently learned concepts, but they recovered quickly during the subsequent mathematics course. Over the longer term, there were no negative effects caused by the gap (Smythe, Stennett, & Rachar, 1974; Stennett & Rachar, 1973). Shockey (1997) conducted a similar study at two suburban United States high schools using semestered block schedules and produced similar results. Participants in her study were 172 students at the two schools who were enrolled in pre-calculus in the spring of 1997. There were three groups of students at each school. One group (n=53) had a "retention interval" of 0 months. These students had completed Algebra 2 during the first semester of the 1996-97 school year, beginning the pre-calculus course within a week of completing Algebra 2. A second group (*n*=55) had a retention interval of 8 months. These students had completed Algebra 2 during the spring of 1996. The third group of students (*n*=64) had a retention interval of 12 months. These students had completed Algebra 2 during the fall semester of the 1995-96 school year.

Students completed district mandated end-of-course tests in Algebra 2 three times: once at the end of their Algebra 2 course, once at the beginning of the pre-calculus course, and once after approximately four weeks during which the pre-calculus teacher reviewed Algebra 2 concepts. In addition, at the end of the pre-calculus course students completed a district mandated end-of-course test in pre-calculus. The Algebra 2 test contained two parts: a 37-item multiple choice test, and a rubric scored performance-based assessment. For the administration at the end of Algebra 2, only scores for the multiple choice portion of the end-of-course assessment were available for analysis, and there was no statistically significant difference on that part of the test. When tested pre-review and post-review during the pre-calculus course, there was no significant difference among the groups on the performance assessment portion of the test.

On the multiple choice portion of the pre-review test taken at the beginning of pre-calculus the students with a retention gap of 0 months significantly outscored the students with a retention gap of 8 or 12 months. At the end of approximately four weeks of review, there was no significant difference between the multiple choice scores of students with 0 months retention gap and those with 8 months' retention gap. Students with 12 months retention gap had partially caught up, but still scored significantly lower than did students with 0 months retention gap. However, when tested on pre-calculus concepts at the end of the pre-calculus course, there was no significant difference among the groups, nor even a trend towards higher scores for the students with 0 months retention gap.

In general, Shockey's (1997) results confirm those reported earlier in London, Ontario (Smythe, Stennett, & Rachar, 1974; Stennett & Rachar, 1973): A retention gap caused students to score more poorly on a test of the immediately preceding course, but had no negative effect on students' ability to learn material in a subsequent course. However, one aspect of Shockey's (1997) data raises a possible concern. She observed a total of four teachers, two in each of two high schools, teaching pre-calculus in a semestered block schedule. One of the teachers spent 15 days reviewing Algebra 2 concepts, one spent 20 days, one spent 21 days, and one spent 22 days. On average, more than 20% of the 90-day pre-calculus course was spent on review.

Would so much review have been necessary had students been enrolled in a regular schedule, so that the retention gap before beginning pre-calculus was 3 months for everyone, instead of ranging between 0 and 12 months? Shockey's data leave this question unanswered. Surveys and interviews have provided mixed messages in response to this same question. Students and teachers at six Ontario schools with semestered block schedules indicated on a questionnaire that students encountered difficulty in returning to a subject after a break of a semester (King, et al., 1978). In contrast, other researchers provided anecdotal evidence that teachers could discern very little difference in the amount of review needed by students who had a retention gap of three months and that needed by students with a longer retention gap (Canady and Rettig, 1995; Kramer, 1997b). In short, this question is still unanswered, and further research will be needed to address it.

Effects of Semestered Block Scheduling on Mathematics Achievement

Student grades. Many studies have analyzed the academic impact of block scheduling by comparing student grades under a block schedule with grades under a traditional schedule. Most have reported that grades under a block schedule are higher (Carroll, 1994; King, et al., 1975; Pulaski County High School, 1995). One case study reported that mathematics grades, in particular, had improved (Reid, 1994).

However, Wild (1998) reported an opposite trend among British Columbia schools in 1995-96 and 1996-97, with students in semestered block schedules tending to have lower teacher-assigned grades than either students in schools with traditional all-year schedules or stunts in schools using "quarter" plans with even longer time blocks. There were several differences between Wild's study and the others, any of which may account for the difference. He compared grades across schools, rather than reporting change of grades within schools; his sample size was larger; and his study was conducted in British Columbia in the mid-1990s.

Whatever the effect of a semestered block schedule on student grades, it is not certain that improved grades reflect increased learning. Three studies have presented data indicating that improved grades under a block schedule may be the result of grade inflation, and thus not a valid measure of academic achievement (Gore, 1995; King, et al., 1975; Wild, 1998).

One study conducted by Cobb, Abate, and Baker (1999) did manage to look at grade point averages while avoiding the problem of grade inflation under a block schedule. They did so by looking at the grade point averages of 150 tenth and eleventh graders in a traditionally scheduled high school who had previously attended a junior high school using a semestered block schedule. They found that this group's grade point averages in tenth and eleventh grade were significantly higher than those of a matched sample of 150 students who had attended one of two other junior high schools in the same geographic quadrant of the city whose school size, ethnic, and socio-economic make-up were comparable. Unfortunately, their study did not disaggregate course grades by subject area, so it is unclear whether the overall higher grade point averages of students who had attended the block scheduled junior high school applied to mathematics grades in particular.

Mathematics test scores: Canadian studies. With the possible exception of the study by Cobb, Abate, and Baker (1999), mathematics test scores are probably more valid than course grades for measuring the impact of semestered block scheduling on achievement. Because semestered block schedules became popular in Canada before they became popular in the United States, most early research on the schedule's achievement effects was conducted there.

Two relatively small scale investigations conducted in London, Ontario found no significant difference between the achievement of students studying under a semestered block schedule and that of students studying under a traditional schedule. As noted above, an early longitudinal study found that a semestered block schedule had no impact on students' mathematics achievement at the end of tenth grade (Smythe, Stennett, & Rachar, 1974; Stennett & Rachar, 1973). A second study compared test scores of 350 students in London, Ontario studying Grade 9 general level mathematics during the second semester under a block schedule to those of 309 students studying the same course under an all-year schedule. Because the test was administered approximately one month before the end of the school year, semestered classrooms had completed an average of only 74.4% of the curriculum, whereas all-year classrooms had completed an average of 82.9% of the curriculum. Nonetheless, scores between the two groups were nearly identical (Stennett, 1985).

Raphael, et al. (1985) performed a much larger scale studying the province of Ontario. They investigated performance on the Second International Mathematics Study

(SIMS) of Ontario students who were in Grade 12 or who were mathematics specialists in Ontario's (college preparatory) Grade 13. "Math specialists" were defined as students taking more than one mathematics course in Grade 13. The SIMS data were collected using a stratified random sample designed to be representative of the Ontario province as a whole. In all, 5280 students from 250 classrooms, 84 of which used a semestered block schedule, participated in the study. Student socioeconomic status was estimated based on student responses to SIMS questionnaire items asking each student's mother's and father's profession. According to this measure, students from classrooms using a semestered block schedule did not differ from students in all-year classrooms in socioeconomic status.

Achievement of students in Grade 12 classrooms using a semestered block schedule was compared to that of students in classrooms using a traditional year-long schedule on eight subscales: Number Systems, Algebra Computation, Algebra Other, Equations and Inequalities, Analytical Geometry, Trigonometry, Functions, and Probability and Statistics. Students in year-long classes outperformed students in semestered classes on all eight subscales, with differences on Number Systems, Algebra Other, Equations and Inequalities, and Analytical Geometry significant at the .05 level. Also, responses of Grade 12 students to an attitude questionnaire found that those in semestered classes generally had less positive attitudes towards mathematics than did their peers in year-long classes (Raphael, et al., 1985).

Achievement of Grade 13 math specialists in classrooms using a semestered block schedule was compared to that of Grade 13 math specialists in classrooms using a traditional year-long schedule on the eight subscales used for Grade 12 students, plus three additional subscales: Complex Numbers, Differentiation, and Integration. Students in year-long classes outperformed students in semestered classes on all eleven subscales. Differences were statistically significant at the .05 level for all subscales except Complex Numbers and Trigonometry.

Despite the careful sampling design, some Canadian educators contacted to provide context for understanding the Raphael, et al. (1985) study indicated that sampling students by mathematics class may have introduced unintentional bias in the analysis. They provided anecdotal reports that in semestered schools lower-ability students in Grades eleven and twelve were encouraged to "try" math courses in the fall, and if they didn't do well to retake the same course in the spring. Such students could have been sampled twice in the scheme used by SIMS. However, a decade later Marshall, Taylor, Bateson, and Brigden (1995) obtained similar results in a British Columbia study that avoided this particular pitfall by testing all tenth graders, regardless of the particular mathematics class in which they were enrolled.

Marshall, et al. (1995) reported results from British Columbia's 1995 Mathematics and Science Assessment. The assessment used a matrix sampling procedure, with each tenth-grade student taking one of four forms of the test. Each form contained 20 questions addressing knowledge covered in the British Columbia mathematics curriculum through tenth grade. The assessment tested 16,356 students who studied Mathematics 10 under a traditional all-year schedule, 6,461 students who studied Mathematics 10 under a semestered block schedule, and 1,703 students who studied Mathematics 10 under a "quarter plan" block schedule. Students who studied under the all-year schedule outscored those who studied under the semestered schedule, who in turn outscored those who studied under the quarter plan. Results were statistically significant. The stability of the differences is perhaps best reflected in the fact that of the 80 items presented among the four forms of the test, Mathematics 10 students in all-year programs scored highest on 74 items, second on 5 items, and lowest on only 1 item. Semester students scored highest on 3 items, second on 68 items, and lowest on 9. Quarter students scored highest on 3 items, second on 7 items, and lowest on 70 items. Marshall, et al. reported similar though less extreme results on the same assessment for all-year, semestered, and quarter students enrolled in Mathematics 10A, a non-college-preparatory version of the Grade 10 British Columbia mathematics curriculum.

Wild (1998) provided further evidence to corroborate the findings reported by Marshall, et al. (1995). He reported participation rates and exam grades for all Grade 12 students enrolled in British Columbia schools in 1996-97. Mathematics 12 is the college preparatory mathematics elective for British Columbia Grade 12 students: nearly all students who wish to take advanced mathematics enroll in the course. Wild reported that during the 1996-97 school year, 51.6% of British Columbia students enrolled in all-year classes completed Mathematics 12 and took the provincial exam, compared to only 34.1% of students attending schools with a semestered block schedule and only 29.1% of students attending schools using a quarter plan. Further, the average marks on the provincial exam were higher for students in the all-year schools: 69.4% in all-year schools, versus 64.6% in schools using a semestered block schedule and 62.5% in schools using a quarter plan. The percent of students in each group who received an "A" on the provincial exam formed an even more striking pattern: 24.3% for all-year students, 14.1% for students using a semestered block schedule, and 10.7% for students using a quarter plan. The distinct pattern cannot be attributed to small sample sizes: in 1996-97 exams were completed by 8,407 students in all-year classes, by 8,936 students in semestered classes, and by 1,163 students in quarter-plan classes. Wild also reported results for 1995-96 that were similar to those for 1996-97.

Lastly, Wild (1998) summarized results of a nationwide study that had been conducted by the Council of Ministers of Education, Canada to investigate mathematics achievement of students at ages 13 and 16. Typically 500-1000 students were sampled in each Canadian province. A total of 12,881 13-year-olds participated in the study, 86% of whom were enrolled in all-year classes and 8% of whom were enrolled in semestered classes. Mathematics achievement was scored on a scale from lowest of 1 to highest of 5. Of 13-year-olds enrolled in all-year classes, 30% scored 3 or higher, versus 24% in semestered classes. A total of 11,079 16-year-olds participated in the study, 40% of whom were enrolled in all-year classes and 49% of whom were enrolled in semestered classes. Among 16-year-olds enrolled in all-year classes, 71% scored 3 or higher, versus 55% in semestered classes.

Taken as a whole, the Canadian studies provide a very clear picture of lower mathematics achievement in classrooms using a semestered block schedule. The one potential weakness shared by all the studies is the possibility of a school-level volunteer effect. Perhaps schools that adopted a semestered block schedule tended to have systematically weaker mathematics achievement before they adopted the schedule change. Raphael et al. (1985) did provide evidence that schools using semestered block schedules were similar in socio-economic to schools using traditional schedules. Further, the current author contacted both Wild and Bateson—one of the Marshal, et al. (1995)

authors—who provided anecdotal evidence that the schools adopting a semestered block schedule were not lower-achieving to begin with. Nonetheless, studies in the southern United States that have systematically investigated this issue have found that schools adopting a semestered block schedule did indeed have lower mathematics achievement before adopting the schedule than did schools which remained on a traditional schedule (North Carolina Department of Public Instruction, 1999; Texas Education Agency Research and Evaluation Division, 1999). The possibility that a similar pattern occurred in Canada cannot be entirely disregarded.

Mathematics test scores: United States studies. In recent years, several case studies have been published comparing mathematics achievement at particular sites in the United States under a semestered block schedule to achievement at the same site under a traditional schedule, but some were of very low quality and are not reviewed here.

Among the higher quality studies, one compared the achievement of students who completed algebra or geometry under a traditional schedule at two high schools in Dothan, Alabama in 1993-94 to that of students who took algebra or geometry from the same teachers under a semestered block schedule in the fall of 1994-95. All students were given a nationally normed standardized test for algebra at the end of the year in May, 1994, or at the end of the fall term in 1995. Although students using the traditional schedule scored higher on both subjects, the differences were not statistically significant (Lockwood, 1995).

Another study compared a sample of 355 students in Grades 8, 9, 10, and 11 who had attended a semestered junior high school to a matched sample of 355 students who attended one of two other junior high schools in the same geographic quadrant of the city whose school size, ethnic, and socio-economic make-up were comparable. Students were paired based on grade level, gender, ethnicity, and fifth grade scores on the Iowa Test of Basic Skills (ITBS). The "semestered block schedule group" of eighth and ninth graders in their study were still using the semestered schedule; the tenth and eleventh graders had graduated from the semestered junior high school and were attending a traditionally scheduled high school. The study found that students who had attended the semestered junior high school scored lower on the mathematics portion of the ITBS than did the matched sample of students who had attended a regularly scheduled high school, but the differences were not significant at the .05 level (Cobb, Abate, & Baker, 1999).

Gruber and Onwuegbuzie (2001) conducted an investigation that closely parallels the current study in overall design. They compared achievement on the Georgia High School Graduation Test (GHSGT) of 146 high school students who graduated from a high school in the state of Georgia in the academic year 1996-97 to achievement on the GHSGT of 115 students who graduated from the same high school in the 1999-2000 academic year. Students who graduated from that particular school in 1997 attended high school on a traditional six-period schedule. In the academic year 1997-1998 the high school adopted a semestered block schedule. Therefore, the 1999-2000 graduating class received instruction via a semestered block schedule for three years. The students receiving three years of instruction under a semestered block scheduled scored significantly lower on the mathematics portion of the GHSGT than did students who had attended high school using only a traditional schedule. The effect size of the difference was .52 standard deviations.

In the United States, three large-scale studies have compared mathematics test scores under a semestered block schedule to test scores under a traditional schedule. Their results are different from those of similarly large scale studies conducted in Canada. Two found little difference between schedule types, and one found a significant difference in Algebra test scores favoring the semestered block schedule.

The Texas Education Agency Research and Evaluation Division (1999) looked at the effects of schedule on student achievement at the 600 of Texas' 1070 high schools for whom complete demographic data was available. The demographic profile for schools on a semestered block schedule was different than that for schools with either alternating-day block schedules or traditional schedules. Semestered schools tended to be in larger districts, to have a larger percentage of ethnic minorities, and to be located in the least wealthy areas of the state. After controlling for these factors, there was no difference by schedule type in percentage of test-takers on campus who passed the Texas Academic Assessment System (TAAS) mathematics test in the spring of 1997, nor on any other academic measure. It should be noted that although the Texas study published both demographic data and conclusions, it did not contain a detailed description of the analysis methodology that produced those conclusions.

A second study analyzed results in Iowa and Illinois on the 1999 administration of the ACT Assessment, a test administered by ACT, Inc. and used by many colleges to assess high school students' general educational development and their ability to complete college-level work. The study reported data for 568 schools, including 351 using traditional all-year schedules, 161 schools using an alternating-day schedule, and 56 schools using a semester plan. After controlling for school demographic characteristics, there were no significant differences by schedule type in any area of achievement on the ACT, including mathematics (Pliska, Harmston, & Hackmann, 2001).

Zhang (2001) reported the latest in a series of studies that have been conducted by the North Carolina Department of Public Instruction to investigate the effects of semestered block scheduling in North Carolina high schools. In North Carolina, the percent of high schools using a block schedule grew from 1.6% in 1992-3, to 35% in 1994-5, to 64.8% in 1996-7, to 73.6% in 1997-8 (North Carolina Department of Public Instruction, 1999). Zhang compared student end-of-course test scores in two groups of schools. One group used a traditional schedule throughout the years 1993-2000, while the other group of schools adopted a semestered block schedule during the peak years of implementation, in 1995, 1996, or 1997. This yielded a sample of 214 Grade 9-12 high schools, consisting of 146 schools who had adopted a semestered block schedule and 68 schools who used a traditional schedule. The data set included scores from 640,000 end-of-course tests in Algebra 1, English 1, Biology, Economic Legal and Political Systems, and US History that had been completed at these schools between the years 1993 and 2000. For purposes of analysis, end-of-course scores were converted to "T-scores," with a mean of 50 and a standard deviation of 10. The reader should note that the term "T-score" as used by Zhang is not related to the *t*-test frequently reported in statistical literature.

Zhang reported the data in two ways, through an Analysis of Covariance (ANCOVA) and by graphically depicting change in student performance over time. The ANCOVA used the mean of school-level scores in 1997, 1998, 1999, and 2000 as a dependent variable and controlled for school-level variables reflecting percent low parent education level, percent free and reduced lunch, percent non-white, and mean school-level subject area score in 1993 and 1994. Because block scheduling was not in wide-spread use in North Carolina in 1993 or 1994, Zhang used the mean of school-level 1993 and 1994 test scores as a covariate. Results of the ANCOVA found significant differences among schools only in Algebra 1 scores. The semestered block schedule schools outscored the traditionally scheduled schools with an adjusted mean of 48.2 points versus 47.2 points.

The effect size of the difference in Algebra 1 scores was small. Since T-scores have a standard deviation of 10, a 1-point difference translates into an effect size of 0.1. However, the ANCOVA obscures a more impressive trend visible in the descriptive data reported by Zhang. Since 1997, algebra scores at the schools using a semestered block schedule have been steadily rising, while those at schools using a traditional schedule have been falling. Although the semestered schools had generally lower Algebra 1 scores through 1997, by 2000 the semestered schools were outscoring the traditionally scheduled schools by approximately two Algebra 1 T-score points. It should also be noted that while semestered schools had a smaller percentage of minority students (in 2000, 34.4% versus 42.2% for traditional schools), they had lower scores on the two measures of socio-economic status. In 2000, the semestered schools had 30% of their students on free or reduced lunch, as opposed to 22% on free or reduced lunch for the traditionally scheduled schools. North Carolina's semestered schools had 49.2% of their students with a parent education level of high school or lower, as compared to 35.2% for traditionally scheduled schools.

Interaction of Semestered Block Schedule with Curriculum

As noted previously, researchers investigating block scheduling have often claimed that in order to be successful it is crucial to modify the mathematics curriculum. Generally, this claim has been based on surveys (King, et al., 1975; Sessoms, 1995) or on the testimony of teachers (Harter, 1994; Kramer, 1996).

The literature on achievement under a semestered block schedule generally supports this claim. Semestered block schedules have correlated with distinctly lower mathematics test scores across Ontario (Raphael, et al., 1985), across British Columbia (Marshall, et al., 1995; Wild, 1998) and in a case study reported in Georgia (Gruber and Onwuegbuzie, 2001). In the case of British Columbia, a provincially mandated curriculum and testing program prevented schools from changing the curriculum when they adopted a block schedule (Kramer, 1997). In the case of Ontario, the curriculum structure described by Raphael, et al. (1985) is quite similar to that in British Columbia, and there appear to be similar structural impediments to changing the curriculum. In the Georgia case study, the authors noted that there were “no notable changes in the curricula” when the school moved from a traditional schedule to a semestered block schedule. According to a personal communication from the authors (April, 2002), this specifically meant that although more courses were offered yearly under the block schedule, students in general did not take a larger number of mathematics courses than they had under the traditional schedule. Students took the same number of courses, and teachers compressed an unchanged curriculum into the shorter time available per mathematics course under the semestered block schedule.

Semestered block schedules correlated with higher mathematics test scores only in the case of Algebra 1 achievement in North Carolina. As reported by Zhang (2001) North Carolina differed from the other sites in that the algebra curriculum was modified under the semestered schedule. Specifically, under the semestered block schedule some students were able to split their algebra study over two courses, entitled Algebra 1a and Algebra 1b.

Interaction of Semestered Block Schedule with Teaching Methods

As noted previously, researchers reporting surveys and interviews of administrators, teachers and students using block schedules of any type have consistently emphasized the need to reduce lecture and increase the use of other means of teaching mathematics (Howard High School, 1994; King, et al., 1975; King, et al., 1978; Meadows, 1995; O'Neil, 1995; Reid, 1995a; Sturgis, 1995; Kramer, 1996).

At the sites where semestered block schedules correlated with lower mathematics achievement, there is evidence that such changes to teaching methods were limited. Raphael, et al. (1995) reported that teachers in Ontario's semester classes were more likely to use workbooks, individualized materials, and visual materials but the authors commented without further elucidation that "the differences, if real, (were) probably not very large (Raphael, et al., 1995, p. 43)." At the Georgia site described by Gruber and Onwuegbuzie (2001) teachers attended professional development in the summer before adopting the block schedule, but despite numerous staff changes there was no follow-up professional development at any time in the succeeding five years, and no systematic attempt to change teaching methods (personal communication, authors, April, 2002). In British Columbia, there was no allocation of planning time that would enable teachers to change their teaching methods. In fact, British Columbia's schools kept planning time constant after switching to a block schedule by allocating to each teacher a full block-scheduled period for planning during some terms, and no in-school planning time during other terms (Bateson, personal communication, January, 1996; Reid, 1995a).

In North Carolina, where Algebra 1 scores increased under a semestered block schedule, planning time in block scheduled schools increased dramatically from one 50 to 55-minute period daily to one 90-minute period daily (Averett, 1994). Despite this, there may have been little change in teaching methods under the semestered schedule. Averett did report that in North Carolina, teachers used a wide variety of instructional practices in semestered classrooms, such as focusing on problem solving, conducting group discussions, and using performance assessments. However, more recent surveys contradict this impression. The North Carolina Department of Public Instruction (1997) reported instructional practices under the block schedule were mostly traditional, that is, lecture, students working at their desks, and small group work. Assessment practices consisted mainly of traditional paper and pencil tests. A survey of 2,167 North Carolina high school teachers, 1,036 of whom taught in traditional-scheduled programs and 1,131 of whom taught in block-scheduled programs, produced similar results (Jenkins, Queen, & Algozzine, 2001). There were few differences by schedule type in the use of most instructional strategies, with the exception that teachers in block scheduled classrooms used student coaching/peer tutoring slightly more often and used projects slightly less often. Both groups of teachers reported more extensive training in direct instruction/lecture than in any other teaching method.

Other authors have reported that under a semestered block schedule there was little change in methods of teaching mathematics (Shockey, 1997) or in teaching methods generally (Shortt & Thayer, 1997). This observation should not be surprising. Making appropriate adaptations to mathematics instruction under a block schedule will involve more than reducing the amount of teacher lecture and relying more on cooperative group work, individual projects, and peer tutoring. As Burrill (1997, p. 3) put it, “You can have students in cooperative groups working on trivial tasks. You can use manipulatives to do rote, meaningless procedures. A teacher can walk around encouraging students but never check their work or their thinking.”

Such implementations of “alternative methods” can have a negative effect on student learning, no matter what the schedule type. As part of the same testing program reported by Marshall, et al. (1995) students in British Columbia completed a survey asking how often various classroom activities occurred in math lessons. The British Columbia Ministry of Education (1995) reported that tenth graders with high scores on the test reported engaging in the following activities more often than did tenth graders with low scores on the test: “The teacher shows us how to do math problems,” “We copy notes from the board,” “We work from worksheets or textbooks on our own,” and “We discuss our completed homework.” In contrast, tenth graders with low scores on the test reported engaging in the following activities more often than did tenth graders with high scores on the test: “We work on math projects,” “We work together in small groups,” and “we check each other’s homework.”

In short, the forms of classroom activity that are most often used to replace lecture are not in and of themselves a panacea. Instead, the key to successfully limiting the amount of lecture is probably to adopt the paradigm advocated by Hiebert, Carpenter, Fennema, Human, Murray, Olivier, and Wearne (1996) and make problem solving the basis for reform in curriculum and instruction. According to these authors, "analyzing the adequacy of methods and searching for better ones are the activities around which teachers (should) build the social and intellectual community of the classroom (Hiebert, et al., 1996, p. 16)." Centering classroom work around these kinds of activities will help to ensure that small group work and mathematics projects engage students in mathematical learning, not just trivial tasks.

At Suburban High School, the vehicle by which the mathematics faculty both redesigned the curriculum to match the timetable available under a semestered block schedule and redesigned instruction to center around mathematical problems was the IMP curriculum. The next section reviews the literature on IMP. It includes a description both of the purposes of the curriculum and of the research to date on achievement effects of IMP and similar curricula designed to implement the NCTM *Curriculum and Evaluation Standards* (NCTM, 1989).

The Interactive Mathematics Program (IMP)

IMP was written by a team of four authors, two of whom had earned doctoral degrees in mathematics and two of whom are classroom teachers of secondary mathematics (Fendel, Resek, Alper & Fraser, 1997). These authors began working in 1989 under a grant from the California Postsecondary Education Commission, and completed the curriculum with the assistance of funding from the National Science Foundation.

Although the authors began their work before the NCTM *Curriculum and Evaluation Standards* (NCTM, 1989) were published, early drafts of the *Standards* were available and from the beginning IMP was designed to comply with them. Thus, the language of the original California grant mirrored many of the key themes that were emphasized in the soon to be published *Standards*. Specifically, the IMP authors were engaged in developing a core high school curriculum that would replace the traditional Algebra I-Geometry-Algebra II/Trigonometry sequence and would set "problem-solving, reasoning, and communication as major goals; include such areas as statistics, probability, and discrete mathematics; and make important use of technology" (Key Curriculum Press, 1998, p. Section II, p. 9). Other goals adopted by the authors were to integrate algebra and geometry and to create a curriculum that would contain numerous small group and individual investigations. Further, in response to the *Standards'* call to develop "mathematical power" in all students, the IMP authors designed a curriculum that could be used by students with a wide range of abilities in a heterogeneous classroom. In order to accomplish these goals, the authors explicitly acknowledged that they would follow NCTM's (1989) advice to de-emphasize paper-and-pencil skills (Alper, Fendel, Fraser, & Resek, 1997).

The IMP curriculum consists of four textbooks, each containing five modular units. Each unit of the IMP curriculum generally begins with a central problem or theme. The problems are generally too complex for students to solve initially. Teachers guide students through a variety of smaller problems that develop the skills and concepts needed to solve the overarching unit problem. Numerous other long-term problems, called "Problems of the Week" are included in each unit (Key Curriculum Press, 1998).

Supplemental material is available in each unit to be used for additional practice or for extension, depending on student needs.

IMP Synergies With a Semestered Block Schedule

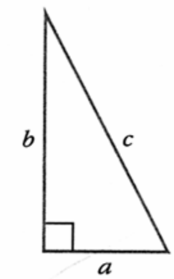
The characteristic of the IMP curriculum that first attracted the interest of the faculty at Suburban High School was its integration of algebra, geometry, trigonometry, probability, and statistics. Mrs. Sullivan, who chaired Suburban High School mathematics department at the time IMP was adopted, reported that she had long been concerned about the her students' lack of ability to use their algebra knowledge when they took her calculus classes. Generally, calculus students had taken Algebra 1 in eighth grade and Algebra 2 in tenth grade. She felt these students tended to "compartmentalize" their algebra knowledge into these two courses and were not fully prepared to apply it in calculus. A block schedule, with potentially longer gaps of time between classes, was likely to make this situation worse. In order to spread algebra study over more courses and better connect it to other areas of mathematics, the Suburban High School faculty decided to integrate the algebra, geometry, and trigonometry curricula when their school adopted a semestered block schedule. It was during discussions with nearby college faculty about the feasibility of writing their own integrated curriculum that Suburban High School teachers were first made aware of the possibility of utilizing the IMP curriculum.

A second aspect of the IMP curriculum that recommends it for use with a semestered block schedule is the active nature of IMP classrooms. According to the curriculum authors, maintaining an active classroom consistent with IMP's philosophy often means "replacing a teacher-led, whole-class discussion with a small-group activity that provides more immediate engagement for students (Alper, et al., 1998, p. 163)." To illustrate the way IMP accomplishes this, the authors presented the "Proof by Rugs" activity shown in Figure 1. After they work on the problem in small groups, some students present their ideas on why the diagrams constitute a proof. The IMP authors have found that when this occurs, all of the students are engaged in the problem and ready to listen to the reasoning. This is precisely the type of activity that the literature reviewed in the previous section indicates may need to be emphasized in order to improve mathematics achievement under a block schedule.

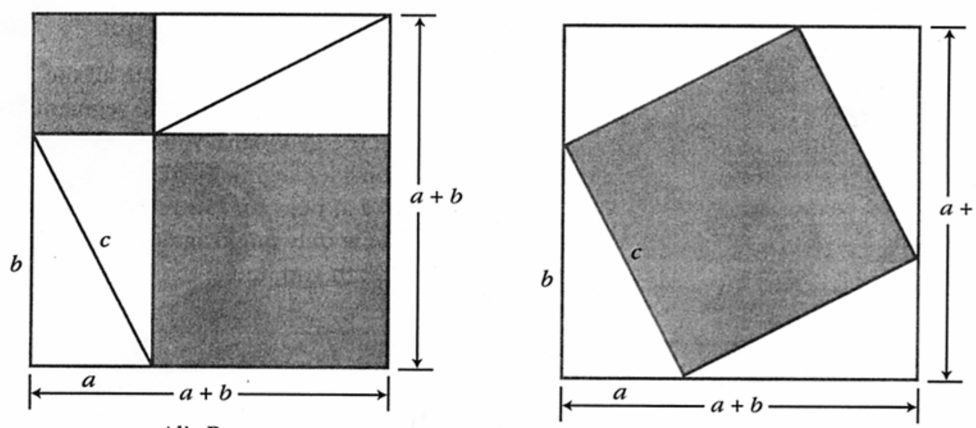
Figure 1. Sample IMP small group activity.

Proof by Rugs

Al and Betty have another game. They began with this right triangle, which has legs of lengths a and b and a hypotenuse of length c. Then they made the two square rugs shown below. Each rug has sides of lengths $a + b$, and the triangles within each square are the same as the single right triangle shown at the right.



When it's Al's turn, a dart drops on the square rug on the left. If it hits the shaded area, he wins a point. When it's Betty's turn, the dart falls on the square rug on the right. If it hits the shaded area, she wins a point. Assume that the darts always hit the rugs, but that they land randomly within the rug. In other words, all points on a rug have the same chance of being hit.



Al's Rug   Betty's Rug

1. Is this a fair game? That is, is the chance of the dart landing on the shaded area the same for the two rugs? Explain your answer.
2. How do the two rugs demonstrate that the Pythagorean theorem holds true in general?

Alper, et al., 1998, p. 164

A related synergy between the IMP curriculum and a semestered block schedule involves planning and professional development. As stated previously, researchers have recommended that teachers be given both opportunities for professional development and additional planning time when adopting a block schedule (Brophy, 1978; King, et al., 1978; Watts & Castle, 1992; Averett, 1994; Meadows, 1995; Reid, 1995a; Canady & Rettig, 1995; Salvaterra & Adams, 1995; Kramer, 1997a). Precisely the same recommendation is made for teachers beginning to teach the IMP curriculum (Key Curriculum Press, 1978). Given the degree to which teaching the IMP curriculum naturally emphasizes the kinds of instruction that work well under a block schedule, learning to implement the IMP curriculum effectively can serve as the focus of professional development and planning time intended to improve mathematics instruction under a block schedule.

A final synergy between the IMP curriculum and the semestered block schedule may be the most important of all. Because the curriculum is composed of 20 distinct modules, it is relatively easy to match the course content to the time available per course under a block schedule, by distributing the modules appropriately among several consecutive mathematics courses. Thus, it is possible to take advantage of the schedule's structure by designing a curriculum in which content is covered over a larger number of courses, with less material contained in each individual course. As noted in a previous section, there are strong suggestions in the achievement literature that absent such an adjustment mathematics achievement is likely to decline under a semestered block schedule.

IMP Connections to Learning Theory

IMP's active and problem-centered approach to instruction is based on a constructivist theory of how students learn (Alper, et al. 1997). The IMP authors quote the following passage from the *Curriculum and Evaluation Standards* to provide a rationale for the active IMP classroom:

> In many classrooms, learning is conceived of as a process in which students passively absorb information, storing it in easily retrievable fragments as a result of repeated practice and reinforcement. Research findings from psychology indicate that learning does not occur by passive absorption alone….Instead, in many situations individuals approach a new task with prior knowledge, assimilate new information and construct their own meanings….This constructive, active view of the learning process must be reflected in the way much of mathematics is taught….Our ideas about problem situations and learning are reflected in the verbs we use to describe student actions (e.g., to investigate, to formulate, to find, to verify) throughout the *Standards* (NCTM, 1989, p. 10).

A second theory that appears to have influenced the development of IMP is situated cognition (Brown, Collins, & Duguid, 1989; Goldman, Petrosino, & the Cognition and Technology Group at Vanderbilt, 1999). Situated cognition is based largely on two lines of research, namely investigations of transfer and investigations of expertise. Investigations of transfer have found that knowledge learned abstractly, in the absence of the context in which it is to be applied, is often "inert;" that is, when students need knowledge that they appear to have learned, they are unable to access it (Bransford, Franks, Vye, & Sherwood, 1989). Investigations of expertise have found that experts tend to have rich, context-dependent schema for understanding and solving problems in their domain of expertise. Often, they develop this expertise by participating in a "community of practice" to obtain objectives of value. Expert tent makers often develop their knowledge as apprentices to more skilled tentmakers; expert mathematicians develop their knowledge working as graduate students with faculty mathematicians. Based on this perspective, advocates of situated cognition advocate having students work in groups on authentic problems that often take extended periods to solve. The influence of situated cognition on IMP can be seen in the curriculum's use of extended problems and on the emphasis placed on cooperative groups. The authors intended for students to "explore open-ended situations actively, in a way that resembles the inquiry method used by mathematicians and scientists in their work" (Key Curriculum Press, 1998, section 2, p. 2).

Critiques of the IMP Curriculum

The IMP curriculum has generated considerable passion and controversy. In 1999, IMP was one of five middle and high school mathematics programs rated as "exemplary" by the U. S. Secretary of Education. Exemplary programs were selected by a three-tiered review, beginning with a review of each curriculum by four-person teams of teachers and others with expertise in education, followed by a similar review of data submitted to document each curriculum's effectiveness, and culminating in a final selection of exemplary and promising programs by a panel of 14 educators, scientists, and policymakers. According to the panel's report, "exemplary programs must be highly rated on quality, usefulness to others and educational significance and must provide *convincing* evidence of effectiveness in *multiple* sites with *multiple* populations (U. S. Department of Education, 1999, p. 1. Italics in the original)."

Soon after the report was released, a group of more than 200 mathematicians, educators, and scientists wrote an open letter to Secretary of Education Richard Riley, protesting the report (Klein, Askey, Milgram, Wu, Scharlemann, & Tsang, 1999). The letter, published as a full page advertisement in the *Washington Post*, criticized the panel for not including active research mathematicians. It further noted that one panel member had published an article claiming that teaching multidigit computational algorithms was a counterproductive and downright dangerous practice—in sharp contrast to a report published by the American Mathematical Society, which noted both the practical and theoretical importance of such algorithms. The letter noted that each of the letter's authors had publicly criticized one or more of the exemplary programs. Finally, it requested that the Secretary withdraw the entire list of exemplary mathematics curricula for further consideration, and urged the secretary to include well-respected mathematicians in any future evaluation.

Wu's detailed review and critique of the IMP curriculum. The debate about the worth of curricula designed to implement the *Curriculum and Evaluation Standards* (NCTM, 1999) has continued, both in journals (Klein, 2000; Fey, 2000) and online. Among the numerous critical articles written, Wu's review of the IMP curriculum (Wu, 2000), which was cited in the original open letter to Secretary Riley, stands out as providing the most detailed critique of the IMP program. In the context of a case study designed to evaluate achievement effects of at a school using IMP, it is worthwhile to list Wu's concerns in detail.

Wu is a mathematics professor who teaches calculus at the University of California, Berkeley. He reviewed five IMP modules in 1992 and updated his review in March 2000, based on changes that had been made to the published form of the curriculum. Wu argued that the IMP curriculum needed to be judged on two criteria. These were its suitability for the approximately 15% of high school students who might consider pursuing a college degree in mathematics, science, or engineering, and its suitability for the remaining 85% who would pursue a different career or course of study.

For the 85% of students who either do not go to college or will not pursue scientific studies in college, Wu concluded that while IMP had serious flaws, he knew of no textbook series that was clearly superior and many series that were substantially worse. Thus, he recommended that for this group of students all teachers would do well to consult IMP often for supplementary materials to be used in the classroom.

Wu's criticisms of the IMP curriculum for the non-mathematical, non-science-oriented 85% of students were based on the perception that if some of these students should change their minds and subsequently pursue a mathematics or science program of study, they should have sufficient knowledge to do so. He listed five specific criticisms.

First, while he felt the curriculum promote understanding, he noted that technical fluency was also important, viewing it as key to mastering the language of mathematics. For this reason, he believed the IMP curriculum should spend more time on drills.

Second, he criticized IMP for failing to follow through on major mathematical ideas by presenting the summary formulas that can be derived from them. It should be noted, however, that the specific examples on which he based his critique were IMP's treatment of derivatives, which was contained in a module that has been replaced in the current version, and IMP's failure to present the quadratic formula, which has been changed in the current version. The quadratic formula is discussed in the current version of IMP.

Third, Wu criticized IMP for abusing open-ended problems and de-emphasizing correct answers. He noted that many of the assignments were designed in such a way that almost any answer could be acceptable, leaving students no standard by which to judge good work from bad.

Fourth, Wu criticized the inclusion of too many mathematical puzzles that could reinforce the misconception that mathematics is nothing but a bag of cute tricks. He was particularly critical of the inclusion of such puzzles on tests, because these items were not likely to test whether a student had learned specific mathematics content well, but whether a student had happened to be inspired at the particular time encompassed by administration of the exam.

Finally, Wu criticized IMP for refusing to acknowledge that mathematics could be inspired by abstract considerations. He felt that there was an almost exclusive emphasis on real world problems. Historically, concepts like negative numbers and complex numbers have often been invented merely to satisfy the internal consistency of a mathematical system, and he felt that implying that all mathematics comes from real world applications would lead to a very biased view of the subject.

In contrast to his analysis for non-specialists in mathematics and science, Wu felt that IMP was far from adequate for the 15% of students considering further study in those fields. He had four primary criticisms.

First, he felt that IMP did not go far enough in abstracting key mathematical ideas. The presentation, he felt, stayed too close to the immediate problem situation and did not then extend ideas to their general applications and inter-connections across many situations.

Second, he felt that the mathematics and science specialists needed more technical drills, just as the non-specialists did. It is interesting to note that his concern for ensuring students develop operational fluency foreshadowed an objective that has been included in the NCTM's new *Principles and Standards for School Mathematics* (NCTM, 2000), but was missing from the *Curriculum and Evaluation Standards* (NCTM, 1989) which helped inspire IMP.

Third, Wu felt that the IMP curriculum lacked sufficient emphasis on precision. He felt the exposition was sometimes so chatty and informal as to lead to sloppiness. He felt that while there were some excellent discussions of proof, these were few and scattered. More importantly, students had insufficient opportunity to see models of rigorous proofs or to write rigorous proofs themselves. Also, some of the open-ended problems were designed in such a way that they could lead potential mathematics or science specialists to believe that an incomplete solution could be an acceptable solution to mathematical problems.

Finally, Wu felt that group activities were over-emphasized. He felt that too little attention was given to individual reflection on the mathematics.

Critiques of constructivism and situated cognition. Much of the criticism of IMP and similar curricula has been based on doubts about the learning theories underlying their design. Anderson, Reder, and Simon (1996) challenged some of the educational recommendations that have been made by advocates of constructivism and situated cognition. These authors did not attempt to debunk either constructivism or situated cognition, noting that under some interpretations they themselves were "constructivists" and had been called so by mathematics educators. Further, one of the authors had previously written a review of situated cognition supporting its compatibility with modern information processing theory. It was the contention of the authors, however, that some of the more extreme proponents of both constructivism and of situated cognition had taken the theories too far, making claims that were contrary to evidence available from research in cognitive psychology. Among the issues raised by Anderson, et al. (1996), this review will address four that are particularly relevant to the IMP curriculum.

The first issue applies directly to the concern of Wu (2000) that there is too little drill in IMP and similar curricula. According to Anderson and his colleagues some constructivists and some advocates of situated cognition have claimed that cognitive tasks cannot and/or should not be decomposed into smaller subtasks. Anderson, et al. presented a large body of evidence demonstrating that cognitive tasks can indeed be broken down into subtasks, and that these subtasks can often be practiced independently of the larger task with fruitful results. They noted a related claim sometimes made by constructivists, that excessive practice or "drill and kill" could lead to routinization and of knowledge and drive out understanding.

This constellation of ideas may have influenced IMP, which according to Wu (2000) in its draft form contained almost no drill and in its published version contains less drill than most traditional texts. However, practice of cognitive subtasks decontextualized from their original context may be critical for making retrieval of those subtasks fluent or automatic. As described in Bransford, Brown, and Cocking (1999, p. 22), within the overall process of solving a problem there are a number of sub-processes that, for the expert, are fluent or automatic. Fluency is important because effortless processing places fewer demands on conscious attention. Since a person can attend to only a limited amount of information at one time, ease of processing some aspects of a task gives the person more capacity to attend to other aspects of the task.

The second issue raised by Anderson, et al. (1996) is the claim by some advocates of situated instruction that abstraction is of little use and that real learning occurs only in "authentic" situations. This idea is closely related to claims that learning seldom transfers between contexts. Because current performance will be facilitated to the degree that the context matches prior experience, the claim is made that the most effective training is apprenticeship to others in the performance situation; abstract instruction, in contrast, is viewed as relatively useless. Anderson and colleagues respond to this claim with evidence from a number of studies that have demonstrated the usefulness of abstraction.

It is clear that the authors of IMP had no intention of avoiding abstraction. As described by Alper, et al. (1997), they often used contexts primarily for motivational reasons, that is, to make the situation concrete enough for students to begin thinking about the problem. But the intent was to start with more concrete situations and build from there to the relevant abstractions. Once students get involved in IMP problems, the authors claim that it is the mathematics, not the context, that holds their attention.

Nonetheless, Wu (2000) claimed that IMP does not go far enough in its use of abstraction. It is possible that IMP, hailing in part from the tradition of situated cognition, was influenced by that philosophy to pursue abstractions less fully than would otherwise be the case. In the end, the question of whether or not the amount of abstraction in the IMP curriculum is sufficient is probably best addressed not by a review of curriculum content, but rather by studies like the current one, which investigate whether students who have utilized IMP are able to apply the knowledge they have learned so that they are successful in future courses.

A third issue raised by Anderson, et al. (1996) is the contention by some advocates of situated cognition that instruction needs to be done in a highly social environment. This is based on the ideas that (a) virtually all jobs are highly social in nature and (b) learning is closely associated with its context. Anderson and colleagues countered with evidence that research on cooperative learning has provided mixed results. They noted that, while useful, cooperative learning is not a panacea. Some learning, in particular drill to fluency in important subtasks, may be best accomplished in individualistic contexts. Both Wu (2000) and Alper, et al. (1997) have noted the extensive use of cooperative learning in IMP classrooms. Wu viewed this as a weakness; Alper and colleagues viewed it as a strength.

This author's view is that cooperative learning can be a powerful vehicle for improving understanding of mathematics, but only if it is introduced from the perspective of constructivism in addition to or instead of the perspective of situated instruction. The key to success is ensuring that mathematics is made "problematic" in the sense described by Hiebert, et al. (1996).

> Allowing the subject to be problematic means allowing students to wonder why things are, to inquire, to search for solutions, and to resolve incongruities. It means that both curriculum and instruction should begin with problems, dilemmas, and questions for students. We do not use "problematic" to mean that students should become frustrated and find the subject overly difficult. Rather, we use "problematic" in the sense that students should be allowed to problematize what they study, to define problems that elicit their curiosities and sense-making skills (Hiebert, et al., 1996, p. 12).

Norms must be developed so that cooperative groups become the locus of debate about different approaches to and understandings of mathematics problems. Student groups must inquire, search for solutions, and resolve incongruities. Groups must analyze the adequacy of student methods and search for better ones. Further the class as a whole must compare and contrast the adequacy of methods devised by different groups. Disagreement among students in the group, and contrasts between the approaches of different groups to the same problem, provide opportunities for the cognitive dissonance that Piagetian theory describes as being key to students making major advances (accommodations) in their understanding.

IMP provides a context in which cooperative groups can be used in this way, but it does not guarantee that they will be. Brombacher (1997) observed five teachers from three cities in the United States who were utilizing the IMP curriculum during its pilot phase. Teacher experience ranged from 4 to 14 years. One of the teachers was teaching his first IMP course; two teachers had taught both IMP 1 and IMP 2; one teacher had taught IMP 1, 2, and 3; and one of the observed teachers had taught all four IMP courses. The teachers Brombacher observed had volunteered to teach IMP and, in some cases, had requested permission to implement the program. In addition to describing many positive aspects of the curriculum, Brombacher listed a number of concerns, including the following:

> In the classes that I watched, all the students sat in groups, but in only a few did I watch students working together on tasks. In the rest, the groups seemed to be forums for general discussion while the teacher was with the other groups. Although I certainly saw teachers trying to "restrain themselves" and in most cases with great success, I have to ask where the student debate was. I never saw any students engage in debate over mathematics. I did see some point out errors to others, but I had hoped to see students really wrestling with ideas and problems, reaching some common solution based on mathematical argument. Sadly I did not. (Brumbacher, 1997, pp. 103-104).

The fourth issue raised by Anderson, et al. (1996) is the claim by some constructivists that since all knowledge is constructed by the learner, direct instruction by the teacher is not a good way to assist student learning. Anderson and colleagues claimed, in contrast, that while in some circumstances people are better at remembering information that they create for themselves, there is considerable research showing they can also remember what they are told. Anderson, et al. (1996) were concerned with the tendency of constructivism to devolve into pure "discovery learning" and noted that investigations of discovery learning have generally produced equivocal or negative findings. According to Anderson and his colleagues, when students cannot construct the knowledge for themselves, they need some instruction. In a similar vein, in their article that advocated enabling children to construct knowledge via problematizing the mathematics curriculum, Hiebert et al. (1996, p. 16) stated, "Our position is that the teacher is free, and obligated, to share relevant information with students as long as it does not prevent students from problematizing the subject."

Two articles by researchers from Vanderbilt University provide a good perspective on the place of instruction in constructivist thinking today. The article titles evoke the main ideas: *New approaches to instruction: Because wisdom can't be told* (Bransford, et al., 1989) and *A time for telling* (Schwartz & Bransford, 1998). The first article, in claiming that "wisdom can't be told" noted that knowledge obtained by "being told" is frequently inert. People when prompted can frequently tell back what they learned, but they fail to use relevant information in unprompted problem-solving situations. The article reports a number of studies demonstrating that, in contrast, a problem-oriented approach to knowledge acquisition, like that of IMP, can lead to knowledge that is not inert. The second article reported a series of studies in which college undergraduates studied fundamental concepts in cognitive psychology. The studies found that there can be an appropriate "time for telling." One group of students analyzed raw data from psychological experiments that reported what information people remembered in various situations. Students were to look for fundamental patterns and principles that determined what would be remembered. Subsequently, they attended a lecture that organized the patterns they had found into a theoretical framework that uses schema theory and encoding theory to predict what people are likely or unlikely to remember. One week later, students were asked to predict outcomes for a hypothetical experiment that could be interpreted in light of the concepts they had studied. Students who had engaged in the problem solving and discovery task followed by lecture were much more likely to use the concepts spontaneously and successfully than were students who had read a summary of the relevant results of the experiments, followed by lecture. They were also much more likely to use the concepts spontaneously and successfully than were another group who spent extra time on discovery of the concepts, but did not receive the lecture.

In light of the view that often "wisdom can't be told," the problem-solving approach central to IMP can be viewed as a real strength that is likely to enhance students' ability to utilize what they learn. However, there is also a danger that IMP teachers will fail to take advantage of appropriate "times for telling." Wu's (2000) criticisms that IMP fails to follow through on major mathematical ideas and fails to provide sufficient abstraction may reflect the results of missed opportunities for "telling"—opportunities that could potentially be addressed in student texts, in teacher manuals, or even in IMP teacher training, but perhaps have not been. The *Teaching Handbook for the Interactive Mathematics Program* (Greene, 2000) advises teachers that they will often need to "bite their tongues" to avoid robbing students of the "Ah-ha!"

experience by telling them the conventional super-formula that can answer a problem with which they are struggling. It provides advice on questioning strategies to move students towards appropriate discovery. But it does not address finding appropriate times to provide information or finding ways to ensure that students have sufficiently abstracted key concepts. Based on recent ideas expressed by Schwarz and Bransford (1999) and by Hiebert, et al. (1996) this may be a shortcoming of IMP as currently implemented.

Studies of Achievement under IMP and other Reform Mathematics Curricula

Because it is very difficult to conduct educational research that is both experimentally rigorous and externally valid, conclusions about the achievement effects of IMP and similar reform curricula will need to be based on the accumulated evidence provided by a large number of studies. The current study will be part of a growing body of research that is beginning to provide such evidence. According to Schoenfeld (2002), the data available so far seem to support the following findings:

1. On tests of basic skills, there are no significant performance differences between students who learn from traditional or reform curricula.
2. On tests of conceptual understanding and problem solving, students who learn from reform curricula consistently out-perform students who learn from traditional curricula by a wide margin.
3. There is some encouraging evidence that reform curricula can narrow the performance gap between whites and underrepresented minorities.

Schoenfeld's (2002) conclusions were tentative and were based largely on data from studies investigating elementary school curricula. This section of the literature review provides a more detailed account of studies in Schoenfeld's review that provided information about achievement in high school mathematics. In some cases, the current review references the original studies, whereas Schoenfeld referenced a forthcoming book containing the studies (Senk & Thompson, in press). The current review also discusses an important article by McCaffrey, Hamilton, Stecher, Klein, Bugliari, and

Robyn (2001) concerning achievement under IMP that was not included in Schoenfeld's review.

Studies of achievement under IMP's sister curricula. Curricula designed to implement the NCTM *Standards* (NCTM, 1989) were developed in the early- and mid-1990s. It takes three or four years after a multiyear high school curriculum has been implemented before students have completed enough of it for their achievement to be tested. Given this timeframe, the first detailed studies of mathematics achievement under the new high school curricula have only recently begun to be published. The current section reviews two such studies, one addressing the University of Chicago School Mathematics Project (UCSMP) *Advanced Algebra* (Senk, et al., 1993) text, and the other addressing the algebra content in the Core-Plus curriculum (Huntley, et al., 2000).

In 1993-94, Thompson and Senk (2001) evaluated the UCSMP *Advanced Algebra* (Senk, et al., 1993) text. Their study investigated only curriculum, not schedule, effects. Although the authors did not specify, the fact that schedule is not described makes it probable that all schools participating in their study used a traditional all-year schedule.

Advanced Algebra is the fourth in a six-book sequence designed in the late 1980s to improve the Grade 7-12 mathematics curriculum. The full series contains the books *Transition Mathematics*; *Algebra*; *Geometry*; *Advanced Algebra*; *Functions, Statistics, and Trigonometry with Computers*; and *Precalculus and Discrete Mathematics*. Like IMP, the UCSMP program was recognized by the Secretary of Education in 1999, but due to a lesser amount of available achievement data, the project was selected as "promising" rather than "exemplary."

Although originally developed before the *Curriculum and Evaluation Standards* (NCTM, 1989) were promulgated, *Advanced Algebra* was revised to be compatible with the *Standards*. The text is not as problem-centered as IMP, but *Advanced Algebra* spends much more time than do more traditional texts emphasizing applications and multiple representations of algebraic concepts. Before adopting IMP, teachers at Suburban High School had used *Advanced Algebra* to teach a course called Algebra 3/Trig. They viewed the text as an intermediate step between a traditional curriculum and a reform curriculum like IMP.

Thompson and Senk's (2001) experimental methodology was remarkably strong, given the real world constraints usually experienced by educational research. Although the research article was written by the authors of *Advanced Algebra*, to minimize researcher bias the study data were collected and analyzed by an outside evaluator. Four schools from varying regions in the United States were recruited to participate in the study: one from a white middle-class suburb of Atlanta, one from a rural area in transition toward becoming a suburb of Chicago, one from a small semi-rural community in Mississippi, and one from an affluent suburb Philadelphia. At each of the four schools, two teachers each teaching two sections of second-year Algebra agreed to participate in the study.

Although students were not randomly assigned to classes, the study was a true experiment at the teacher level. Within each school, one participating teacher was randomly selected to teach UCSMP *Advanced Algebra*, while the other continued to teach the school's traditional Algebra 2 text. It turned out that there were three "traditional" texts used among the schools studied, representing the three most commonly used Algebra 2 tests at the time the study was conducted. Teachers selected to use the *Advanced Algebra* text received a minimal amount of professional development not received by teachers of the Traditional text, consisting of two one-day meetings in Chicago, one in the fall and one in the spring.

The experimental procedure resulted in eight classrooms using the *Advanced Algebra* curriculum, two in each school and eight classrooms using a traditional text, two in each school. All participating classes were either heterogeneous in schools with no tracking or designated "average" in schools with tracking. Altogether 150 students participated in UCSMP *Advanced Algebra* classrooms, and 156 students participated in control classrooms.

Data was analyzed comparing matched pairs of classrooms. At the beginning of the school year, each UCSMP *Advanced Algebra* classroom was matched to the control classroom in the same school it most closely resembled, based on pretest scores and demographic characteristics. There were no statistically significant differences between the two members of any matched pair on the pretest. Further, students in UCSMP classes resembled those in comparison classes in race and gender.

At the end of the school year, students were administered a UCSMP designed test to assess the core content of second-year algebra. The test contained a 36-item multiple-choice section addressing the topics of linear expressions, equations and inequalities; quadratic expressions, equations, and functions; higher degree polynomials and general properties of functions; powers, roots, exponents, and logarithms; variation; sequences and matrices; and trigonometry. In addition, the test contained an Advanced Algebra "Problem Solving and Understanding" subtest, consisting of six free-response items designed to measure students' abilities to solve multistep problems. The six items were chosen because each was solvable using any of several strategies, including numeric, symbolic, and graphical methods, and each required students to explain their reasoning.

At the end of the year, each teacher was asked which of the questions on the multiple choice test was "fair" to his or her students, in the sense that the content it was testing had been covered in class. Then, each UCSMP classroom was compared to its matched-pair control classroom on a "fair test" comprised of only those items that both the UCSMP *Advanced Algebra* teacher and the matched control teacher reported they had covered during the year. This yielded a unique "fair test" for each experiment/control pair of teachers participating in the study. Thompson and Senk (2001) reported eight comparisons of UCSMP *Advanced Algebra* classrooms to matched control classrooms on these "fair tests". They analyzed this data by performing matched-pairs *t*-test with seven degrees of freedom using classroom as the unit of analysis. This yielded a significant difference in favor of UCSMP, t=3.57, p=.009.

In performing the matched-pairs t-test with seven degrees of freedom, Thompson and Senk (2001) can be criticized for using classroom rather than teacher as unit-of-analysis. Because the *Advanced Algebra* classes may by chance have had better teachers than the comparison classes, a more appropriate method would have been to use a hierarchical linear model with students nested within classrooms, nested within teachers, nested within schools. An alternate approach that while less able to detect differences between UCSMP and non-UCSMP classrooms would preserve the nominal significance level, would be to aggregate student scores to teacher-within-school, and then to perform a t-test on four matched pairs, each pair consisting of a UCSMP teacher and a non-UCSMP teacher within a school. Fortunately, data available within the article made it possible for the current author to perform this latter analysis. The result confirmed the statistically significant results reported by Thompson and Senk, yielding a t-statistic of 3.239 with three degrees of freedom, p=.048.

Looking in detail at differences between paired classrooms on the "fair test", the *Advanced Algebra* class outscored its matched traditional class in seven of the eight comparisons. Differences between paired classrooms were significant for four of the eight pairs, in all cases favoring the UCSMP class.

Thompson and Senk (2001) also compared classrooms to their matched pair on a "Conservative" test, consisting of 15 out of the 36 original multiple-choice items that all eight teachers said they had covered in class. While the original 36-item test had been designed to contain items testing a balance of skills, properties, uses, and representations, the majority of items on the Conservative test measured skills. On the Conservative test a matched-pairs *t*-test with seven degrees of freedom using classroom as the unit of analysis yielded a non-significant difference in favor of UCSMP, $t=1.843$, $p=.108$. As with the analysis of data on the "fair test," the current author ran a second analysis on the "conservative test" reported by Thompson and Senk (2002) using a more conservative statistical procedure with teacher as unit-of-analysis and three degrees of freedom. The results confirmed Thompson and Senk's conclusions, yielding a non-significant difference with a t-statistic of 1.516, $p = .227$.

However, looking in detail at differences between paired classrooms on the "conservative" test, the *Advanced Algebra* class outscored its matched traditional class in six of the eight comparisons. Differences between paired classrooms were significant for three of the eight pairs, in all cases favoring the UCSMP class.

Finally, Thompson and Senk (2001) compared classrooms to their matched pair on the 6-item Problem Solving and Understanding test. While all teachers in UCSMP *Advanced Algebra* classrooms reported that they had covered material on all six items, comparison teachers reported that they had covered between 50% and 83% of the items. Given the difference in opportunity to learn, it is not surprising that for the Problem Solving and Understanding test a matched-pairs t-test with seven degrees of freedom using classroom as the unit of analysis yielded a significant difference in favor of UCSMP, $t=4.97$, $p=.002$. As with the analyses of data on the "fair test" and on the "conservative test," the current author ran a second analysis on problem-solving data reported by Thompson and Senk (2002) using a more conservative statistical procedure with teacher as unit-of-analysis and three degrees of freedom. The results again confirmed Thompson and Senk's conclusions, yielding a t-statistic of 16.951, $p < .0005$.

Looking in detail at differences between paired classrooms, the *Advanced Algebra* class outscored its matched traditional class in seven of the eight comparisons. All seven differences favoring the UCSMP classes were statistically significant, while the one difference favoring a comparison class was not. As can be deduced from the number of significant differences, at least one *Advanced Algebra* classroom showed statistically significant higher performance than its matched pair in at least one class at each of the four schools participating in the study.

Huntley, et al. (2000) reported a study that was conducted in 1997 to evaluate algebra achievement under the Core-Plus Mathematics Project (CPMP). Core-Plus bears a closer resemblance to IMP than does the UCSMP. Its development was funded as part of the same NSF project that funded IMP, and, like IMP, Core-Plus was designed to implement the vision of the *Curriculum and Evaluation Standards* (NCTM, 1989). Like IMP, Core-Plus was among the five mathematics programs that in 1999 was recognized as "exemplary" by the Secretary of Education. The 3-year core curriculum is comprised of 21 connected units comprised of several multi-day units in which major ideas are developed through investigations and applied problems. Like IMP, Core-Plus integrates algebra, geometry, trigonometry, statistics and probability, and linear functions, and makes extensive use of graphing calculators. Probably the most significant difference between the two curricula is that each Core-Plus unit is designed around one overarching mathematical theme, whereas each IMP unit is designed around one central problem.

Topics in the Core-Plus curriculum are organized in a concept-then-skills-then-abstraction order.

Of the 21 units comprising the first three Core-Plus textbooks, seven deal primarily with algebra, while an additional three units apply and extend algebra concepts and skills in the context of studying other mathematical content areas. It was algebra achievement that Huntley and her colleagues set out to investigate. In the Spring of 1997, among 36 schools that were piloting the Core-Plus curriculum the researchers recruited six schools that would participate in the study, two from the Southeast, two in the Midwest, one in the South, and one in the Northwest. At each site, one or more classes completing the third Core-Plus textbook participated in the study, as did one or more comparison classes that were studying the third year of high school mathematics using a more traditional program. Each comparison class, selected from the same school or a nearby school, was to be comparable in ability to the Core-Plus class.

Unlike Thompson and Senk (2001), Huntley, et al. (2000) were unable to implement a random experimental design. Instead, they performed a quasi-experiment on already intact classrooms. At four of the six sites, they used eighth-grade test scores to ensure Core-Plus groups and comparison groups were comparable. At one of these four sites, students had comparable ability on entry into high school. At three others, the researchers used blocking techniques to match CPMP students with comparison-class students who had comparable mathematics achievement or aptitude scores in Grade 8. Of the two sites where Grade 8 test scores were unavailable, one had randomly assigned students to Core-Plus and control treatments on entry into Grade 9. At the remaining site, the researchers relied on repeated assurances from the school that the two groups were indeed equivalent. At some sites both Core-Plus and comparison students were below average in prior achievement, at some sites both groups were above average in prior achievement, and at one site both groups were heterogeneous.

At each site, the researchers collected data over two days in April or May. At the time of testing, the Core-Plus group at five of the six sites had completed all algebra units in the first three textbooks before the time of testing; at the sixth site, where the class contained lower-ability students, they were just beginning the second of three algebra units contained in Course 3. Comparison classes used a wide variety of textbooks, including advanced algebra texts, a discrete mathematics with applications text, a text focusing on mathematics applications, and a text focusing on the use of mathematics in business settings. It should be noted that, since not all the comparison groups utilized algebra texts, some students in the comparison groups may have had limited opportunity to learn algebra content. This provides a possible alternate explanation of results obtained by Huntley, et al. (2000) and should be kept in mind when interpreting those results.

As noted previously, the Core-Plus researchers collected algebra achievement data using an assessment instrument that was also utilized for the current study. Their version used the form for Part 1 contained in Appendix A of this paper, plus three parallel forms. They used the form for Part 2 contained in Appendix A, plus one parallel form, and the form for Part 3 contained in Appendix A, plus two parallel forms. Forms of the assessment were randomly distributed among students at the time of testing.

For each part of the test, Huntley and her colleagues performed a simple comparison of the mean score across all forms of Core-Plus students to the mean score across all forms of Comparison students. On Part 1, Performance on Applied Algebra Problems with Use of Calculators, the Core-Plus students scored higher than the comparison students by about 0.46 standard deviations. Core-Plus students scored higher than comparison students at five of the six research sites. The authors reported a statistically significant difference (t_{560}= 5.69; p<.001). However, the reported t-test used student as unit-of-analysis, which can be criticized because student scores within a curriculum program (Core-Plus or Comparison) at a given site might not be independent from one another. A more conservative procedure would be to aggregate data to the curriculum-within-site level, and perform a matched-pairs *t*-test with five degrees of freedom. The current author did so, and could not confirm statistical significance (t=1.581, p=.175.)

On Part 2, Performance on Algebraic Symbol Manipulation Without Use of Calculators, the Core-Plus students scored lower than the comparison students by about 0.54 standard deviations. On Part 2, Comparison students scored higher than Core-Plus students at all six research sites. The authors reported a statistically significant difference (t_{575}=-6.50, p<.001). As with Part 1, the current author re-analyzed the data by performing a more conservative matched-pairs *t*-test with five degrees of freedom, using program-within-site as unit of analysis. Again, the statistical significance could not be confirmed (t=-2.455, p=.058).

Part 3, Performance on Open-Ended Applied Algebra Problems with Use of Calculators, was completed by students working in pairs. The Core-Plus pairs scored higher than the Comparison pairs, by about 0.28 pair-level standard deviations. On Part 3, Core-Plus students scored higher than comparison at five of the six research sites. The authors reported a statistically significant difference (t_{364}=2.77, p<.01). As with Parts 1 and 2, the current author re-analyzed the data by performing a more conservative matched-pairs *t*-test with five degrees of freedom, using program-within-site as unit of analysis. Again, the statistical significance could not be confirmed (t = .992, p = .367).

Because the more conservative statistical analyses using curriculum-within-site as unit of analysis could not confirm statistical significance, it cannot be certain that the differences reported by Huntley, et al. (2000) were non-chance. It is probably best to view the study by Huntley, et al. (2000) as a case study of an early implementation of the Core-Plus curriculum, which provided some indication that Core-Plus students were better than traditionally educated students at solving the types of algebra problems emphasized by the Core-Plus curriculum, but somewhat less good at solving the types of symbol manipulation problems emphasized by the traditional curricula.

In addition to the summary results for Parts 1, 2, and 3, Huntley, et al. (2000) reported details about how well students in the Core-Plus and Comparison classes performed on specific sub-skills within Part 1 and Part 2 of their test. Core-Plus students were much stronger than Comparison students at formulating algebraic models to describe a problem situation and at interpreting the meaning of an algebraic model presented in a problem situation. Core-Plus students were also better at what Huntley and colleagues called "representational fluency," that is, translating among graphs, tables, and algebraic symbols to represent a function. Students in Comparison classrooms were much stronger than Core-Plus students at performing algebraic calculations without context or calculator access. However, when similar calculation problems were presented in context and calculators were available, the Core-Plus students were slightly stronger than those in Comparison classrooms.

Huntley, et al. (2000) made one last observation about how Core-Plus was implemented. At site 4, teachers supplemented the Core-Plus curriculum with materials that gave students more practice on traditional algebraic skills. Site 4 was the only site at which Core-Plus students matched the performance of control students on Part 2, Algebraic Symbol Manipulation Without Access to Calculators. This observation is interesting in context of the current study, because although teachers at Suburban High School used a largely unmodified version of IMP for students who participated in the Algebra Achievement testing, in later years Suburban High School teachers supplemented IMP with materials similar in concept to those used at site 4 in the Core-Plus study.

Studies of achievement under IMP. Webb (in press) reported a series of studies designed to assess IMP's effects on students who utilized the program during its pilot years. His studies were conducted at nine schools between the years 1993 and 1997. Five of the schools studied were in California, two in the East, one in the Midwest, and one in a mountain state. In the series of studies, Webb addressed three questions. How did students who used IMP differ from comparable students enrolled in a traditional curriculum in the number of college-qualifying mathematics courses they studied during high school? How did students who used IMP differ from comparable students enrolled in a traditional curriculum in standardized test scores, as reflected on their high school transcripts? How did the achievement of students who used IMP differ from that of comparable students enrolled in a traditional curriculum in content areas where IMP has tried to increase emphasis, that is, probability and statistics and complex problem solving?

Webb addressed the question of mathematics course taking by examining the transcripts of 1,121 students who graduated from three high schools in California in 1993. At the time, they were the only schools at which students had had an opportunity to complete three years of IMP, generally in Grades 9 through 11. All three high schools served diverse student populations and offered a full range of mathematics courses, from basic mathematics to Advanced Placement Calculus. All three schools offered students the option of enrolling in IMP or traditional mathematics, so at all three schools the IMP students were volunteers. Webb used prior test scores or course grades to ensure that students in the IMP group demonstrated prior mathematics ability that was basically comparable to that of students taking traditional mathematics courses.

Webb found that 64% of students who began their high school career in ninth grade by taking IMP Year 1 enrolled in four or more years of college-preparatory mathematics during high school, as compared to only 38% of students who began their high school career in Grade 9 by taking Algebra 1. A similar effect was observed in each of four ethnic groups that were represented across the three schools: Asian/Pacific Islander/Filipino, Black, Hispanic, and White. Webb reported that the result was statistically significant ($p<.01$) but did not account for possible within-school correlation when performing the statistical test.. Differences between the two groups in the proportion of students taking at least three years of college preparatory mathematics, or in the proportion of students taking pre-calculus or calculus, were negligible and not statistically significant.

Webb also looked at standardized test scores reported on student transcripts. At each school where data was available, Webb compared SAT mathematics test scores and/or Grade 11 CTBS mathematics achievement test scores of students who took IMP in ninth grade to those of students who took Algebra 1 in ninth grade, and of students who took IMP in tenth grade to those of students who took Algebra 1 in tenth grade. Results varied widely among the three schools, but only one contrast was statistically significant. At one of the schools, students who enrolled in IMP in ninth grade scored significantly higher than did those who enrolled in Algebra 1 in ninth grade on SAT mathematics. This was true despite the fact that a larger percentage of students who enrolled in ninth grade IMP eventually took the SAT (34% for IMP students, versus 26% for Algebra 1 students).

Webb also addressed a question of great concern to parents in the school communities: Were students of particularly high ability likely to be harmed by IMP? To do so, at the school that had Grade 7 CTBS scores available, he created a matched sample of high-ability IMP students and high-ability Traditional students. For each group, Webb selected students who scored at the 76^{th} percentile or higher on the Grade 7 CTBS, and were enrolled in their respective mathematics curriculum (IMP or Traditional) for at least 2.5 years. Fortuitously, this yielded 58 high-ability students in both the IMP and Traditional groups.

Regarding SAT scores, although differences between the two high-ability groups were not statistically significant, a larger percentage of IMP than of Traditional high-ability students took the SAT (83% versus 74%). Further, the IMP students had a higher mean SAT score, 544.8 versus 530.9. The high ability IMP students also had a higher grade point average than did the high ability Traditional students, both in mathematics and in all subjects excluding mathematics. Webb interpreted this to mean that involvement in IMP might be helping students in other courses as well. An equally likely possibility is that the difference in grade point averages indicates that, even though the two groups had similar prior test scores, there was some difference in attitude or ability that both made students more likely to volunteer for IMP and more likely to do better at school.

At the end of the 1995-96 school year, and again at the end of the 1996-97 school year, Webb tested student achievement on content emphasized by the IMP curriculum. He used three measures. In Grade 9, students completed modified versions of four statistics items that had been used by the Second International Mathematics Study (SIMS). In Grade 10, students completed two multi-step open-ended performance assessments prepared for the Wisconsin Student Assessment System. One of the items required some knowledge of probability and the other item required some knowledge of combinatorics. In Grade 11 students completed 10 multiple-choice items from a practice version of a quantitative reasoning test that was used by a prestigious university to screen its first-year students. The ten items focused mainly on data interpretation and sought evidence of how students used mathematics, probability, statistics, and computation to solve problems. Six high schools participated in this part of the study, although not every one of the six participated in every grade-level test. Webb completed all statistical analyses after controlling for eighth-grade test scores.

Grade 9 results appeared to demonstrate that opportunity to learn was the key to doing well on these assessments. At two of the three schools, the IMP students scored significantly higher on the-ninth grade probability and statistics test than did comparison students taking the traditional college-preparatory course (either Algebra 1 or Geometry, depending on the school). At the third school, the first year algebra course had been "enhanced" by the teachers to include a unit on probability and statistics. At that school the Algebra 1 students scored significantly higher than did the IMP students on the ninth grade probability and statistics test. The three effect sizes were +1.31, +.83, and -.76.

Tenth-grade results were similar, though a little more favorable to IMP. IMP students significantly outscored students in traditional classes on the open-ended problem at two of the three schools. Interestingly, one of the two schools where IMP students performed better utilized the UCSMP Algebra 1 and Geometry books for its "traditional" curriculum. At the third school, which had "enhanced" its geometry curriculum by adding a unit on problem solving and combinatorics, achievement of the IMP and Comparison groups was not significantly different, although the IMP students did slightly better. The three effect sizes were +1.04, +.74, and +.08.

The Grade 11 test was completed at two schools. In both cases, the IMP group scored better than did the Traditional group taking Algebra 2. The two effect sizes were +1.15 and +1.24.

Overall, Webb's series of studies seemed to support the conclusion that IMP encouraged students to enroll in a larger number of college-preparatory mathematics courses, IMP did not harm and perhaps helped students on standardized tests, and IMP succeeded at teaching the content that was intentionally emphasized by the curriculum. The studies also demonstrated that other approaches could be equally successful at teaching probability, statistics, and other content emphasized by IMP, so long as teachers chose to give their students opportunities to learn such material.

McCaffrey, et al. (2001) provided an important additional perspective on how IMP can impact mathematics achievement. They looked at the relationship between instructional practices and achievement on the Stanford 9 test in 226 tenth-grade mathematics classrooms located in a large urban district that was pursuing reform under an Urban Systemic Initiative. One hundred eighty seven of the classrooms taught traditional algebra or geometry, while the remaining 39 used either IMP or a similar curriculum called College Preparatory Mathematics (CPM). The authors found that increased use of reform teaching practices consistent with inquiry-based instruction, as measured by teacher responses to 17 survey items that asked how frequently they engaged in various activities consistent with inquiry-based instruction, predicted better achievement in classes using IMP or CPM, but was unrelated to achievement in classes using more traditional curricula.

The study by McCaffrey, et al. (2000) makes an interesting contrast to data provided by the British Columbia Ministry of Education (1995). As discussed earlier, that study found that tenth graders with low scores on a provincially administered mathematics assessment were more likely than students with high scores to report that "We work on math projects" and "We work together in small groups." In contrast, the study conducted by McCaffrey, et al. found seemingly contradictory results. Their measure of "reform teaching practices," was partly made up of questions that measured how frequently teachers made use of the same activities that had correlated with lower achievement in British Columbia. In contrast to the British Columbia results, McCaffrey and her colleagues found that "reform teaching practices" correlated positively with achievement in IMP classes.

It is possible that either the report by the British Columbia Ministry of Education (1995) or the report by McCaffrey, et al. (2000), or both, contained invalid or misleading results. The British Columbia study reported which classroom activities occurred "more frequently" in the classes of lower-achieving students, but provided no data on how "lower achieving" and "higher achieving" students were defined, on precisely how much "more frequently" the activities occurred, or on statistical significance of the findings. Further, the British Columbia report did not control for prior student ability in determining its results. The study by McCaffrey and her colleagues provided considerable detail on how their measure was defined and the magnitude of results, as well as a careful statistical analysis. The authors did control for prior ability. However, their measure of "reform teaching practices" may not have validly captured the construct they intended it to. For example, a teacher ostensibly using IMP who reports that his or her students "rarely" work on extended mathematics investigations or projects is probably not in fact utilizing the intended curriculum. In IMP classes, a low score on the "reform teaching practices" scale could in part be measuring lack of compliance with the course syllabus, rather than low implementation of inquiry based instruction.

Despite these concerns, it seems likely that both studies do in fact capture an aspect of reality. Probably, teachers in British Columbia engaging in small group work and extended projects were less successful than others in preparing their tenth-grade students for the provincially administered mathematics assessment. Probably, among teachers using IMP in the urban district studied by McCaffrey and her colleagues, teachers using these same instructional practices were more successful than others in preparing their tenth-grade students for the mathematics portion of the Stanford 9 test. As noted previously, the key difference may be the degree to which the "reform based" activities were used as a means to ensure students problematized the mathematics they studied (Hiebert, et al., 1996). The extended problems in IMP are designed to engage students in deep thinking about mathematics. Further, IMP's problem-centered structure is likely to make it easier for teachers to use group work as a means to engage students in deep thought about mathematics, rather than as an end in itself.

As noted previously, adopting IMP makes it easier to adapt mathematics instruction in two ways that may be key to success under a block schedule: modifying the curriculum so that the amount of content covered in each course fits the schedule appropriately, and adopting a more reform oriented, inquiry based approach to teaching.

The study by McCaffrey, et al. (2000) provides evidence that, in addition, reform practices adopted to accommodate the block schedule are more likely to have a positive impact on student achievement in classrooms utilizing the IMP curriculum than in other classrooms.

In short, there is reason to believe that the IMP curriculum not only makes it easier to adapt mathematics instruction in ways that fit well with a semestered block schedule, but further the IMP curriculum makes it more likely these adaptations, once they have been made, will be successful in improving student achievement. In theory, then, one might expect to see a particularly positive impact on student mathematics learning at a school that adopts a semestered bock schedule and at the same time adopts the IMP curriculum. The current study was designed to test this theory.