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 76th 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.