

***ASSESSING THE COSTS/BENEFITS OF AN NSF
“STANDARDS-BASED” SECONDARY MATHEMATICS
CURRICULUM ON STUDENT ACHIEVEMENT***

*The Philadelphia Experience:
Implementing the
Interactive Mathematics Program (IMP)*

PART 1: 1993-94 to 1997-98

By

F. Joseph Merlino

Project Director

&

Edward Wolff

Co-Director

The Greater Philadelphia Secondary Mathematics Project

April 2001

Acknowledgments

This report was made possible by grants from the National Science Foundation. The authors of this report wish to thank:

- Ms. Clara Tolbert, former Director of the Philadelphia Urban Systemic Initiative, for all her years of collaboration and support for this project.
- Dr. Alice Jordan for her seven years as co-director of the project.
- Drs. Tom Clark, Robert Offenberg, and Lori Westler, from the School District of Philadelphia, for their support in collecting and analyzing the data.
- All the IMP teachers whose hard work and dedication made the success of this project possible.
- Drs. Dan Fendel, Diane Resek, Sherry Fraser and Lynne Alper, the authors of the Interactive Mathematics Program.

Executive Summary

Cost/Benefit Question:

All things being equal, does an NSF sponsored standards-based mathematics curriculum, such as the *Interactive Mathematics Program (IMP)*, have a greater positive impact on student achievement than a “traditional” pre-standards program to such a degree as to justify the time, energy and cost of implementing it?

Conclusions:

Student Achievement. After nearly five years of collecting student achievement data in the Philadelphia public schools related to the [*Interactive Mathematics Program \(IMP\)*](#), the results boil down to this:

When IMP students were taught by teachers who had been properly trained, IMP students consistently outperformed similar students who were taught using a pre-NCTM standards curriculum and subjected to lecture style instruction. The superior performance results of IMP were found using a variety of measures and across different student ability levels, when measures for achievement are controlled for 8th grade cohort effects. That is, lower ability IMP students did better than their lower ability counterparts while higher ability IMP students did better than their higher ability counterparts.

Cost: The cost to properly implement IMP is approximately **21,000 dollars per teacher over four years**. This includes new textbooks, a classroom set of graphics calculators, classroom materials, an overhead projector and LCD panel, 240 hours of training per teacher, 50 hours of classroom mentoring, and on-going administrative support.

Why Change?

The past twenty years have witnessed a growing and intensified public demand to raise academic achievement for all students. This demand reflects the historic shifts from an agricultural-based economy and society in the 19th century, to an industrial economy and urbanized society in the early and mid 20th century, to a knowledge-based global economy and data driven digital society in the late 20th and 21st centuries.

Mathematics has been valued for its applications in national defense, industrial processes, financial management, medicine and all of the social sciences. For these reasons, student achievement in mathematics, along with English language arts, has been used as one indicator of the general health of schools as well as of the nation's general intellectual capacity. Periodically, startling statistics and events have awakened the public's interest in students' mathematics competence. For example, in 1941, at the beginning of America's involvement in World War II, in a test of 4,200 candidates for naval officer, 62% failed math reasoning---a critical faculty for navigation at sea. The general population was not much better. In 1950, only 35% of persons over 25 years of age finished four years of high school, and less than 14% of African-Americans did. For many of those students who did manage to graduate high school, only two math courses were commonly required. Those courses often consisted of business math, consumer math, or general math, all of which were basically arithmetic at a middle school reading level dressed as a high school text.

Despite massive support for higher education offered by the GI Bill, by the late 1950s less than 8% of the population had completed four years of college, and less than 4% of African-Americans had. (The GI Bill did much more for high school graduation rates than college.) High-level mathematics achievement was reserved for a small elite group. But in 1957, the Soviet Union's launch of Sputnik shocked the country. In the midst of the nuclear-tipped Cold War, fear

swept the nation that America's best and brightest in math and science may not be the world's best and brightest. The First International Mathematics and Science Study conducted during the 1960s documented America's lackluster performance in mathematics and science as compared to other First World countries.

By 1970, 78% of the nation's 25 to 29 year olds had graduated from high school, 43% had completed at least one year of college, and 22% had completed four years of college. The baby boomers were by far the best-educated cohort America had ever seen at the time. But the OPEC oil embargoes and resulting recessions in 1974 and 1978 again jolted the American public to the reality of global competition as Sputnik had a generation earlier. This time, however, the threat was more economic than military. High mathematics achievement could no longer be reserved just for the few. Intellectually skilled workers were needed throughout the emerging knowledge-based economy.

In the midst of the 1982 recession, President Reagan commissioned his Secretary of Education to study the condition of American education in relation to changed global economics. More workers needed higher intellectual skills for the rapidly developing military, industrial, financial, and medical applications and innovations. The result was a landmark 1983 report entitled, *A Nation At Risk*, which warned that American education was adrift in a "rising tide of mediocrity" that was equivalent to "a unilateral act of educational disarmament". The Second International Math and Science Studies (SIMSS) conducted in the mid-1980s confirmed the *Nation At Risk's* assessments. Other subsequent reports throughout the 1980s continued to toll the bell for change as they documented the mathematical deficiencies of American students in skills, understanding and non-routine problem solving. Even the research community was alarmed at America's growing dependence on importing foreign-born research scientists.

By the late 1980s and early 1990s, the mathematics education community responded with a host of reports that called for fundamental changes and upgrades in mathematics education. These reports advocated not only more years and higher levels of mathematics for students to graduate high school, along with better trained and paid teachers, but also a *re-conceptualization* of what constituted important mathematics content, effective teaching practices and authentic assessments¹. These upgraded learning goals required students to acquire a deeper understanding of the core concepts of mathematical and scientific content. Students needed to become comfortable, confident and competent with the processes of scientific inquiry; no longer was it sufficient merely to master mindless robotic mathematical tasks akin to the factory assembly line.

All of these reports delineated an expanded concept of "basic math" found in the elementary grades to include higher-order algebraic and geometric thinking, statistical inference, probability, modeling, and use of calculator and computer technology to solve multi-system problems with large data sets. Moreover, rather than organizing mathematics education into an amalgam of isolated topics, these reports advocated the integration of mathematics topics into fewer core concepts to delve deeply into mathematics' "big ideas." The 1989 National Council of Teachers of Mathematics *Curriculum and Evaluation Standards*, for example, stressed the importance of mathematical reasoning and communication skills for the purpose of fostering "mathematical power," for *all* students. NCTM's *Professional Teaching Standards* have called for changes in the traditional "teacher-telling/student-listening" teaching paradigm. Teachers are urged to lecture less and facilitate more inquiry-based learning activities. By stimulating students' active

¹These reports include: the [National Council of Teachers of Mathematics' \(NCTM\), *Curriculum and Evaluation Standards for School Mathematics* \(1989\)](#); the National Research Council's (NRC) *Everybody Counts: A Report to the Nation on the Future of Mathematics Education* (1989); NRC's *Reshaping School Mathematics: A Philosophy and Framework for Curriculum* (1990); NRC's *Summit on Mathematics Assessment* (1991); and NCTM's *Professional Standards for Teaching Mathematics* (1991).

experimentation with engaging mathematical problems in interesting contexts, teachers can instill in students a deeper understanding of mathematics. Accordingly, assessment of a student's mathematical knowledge must extend beyond computational proficiency to include non-routine problem solving embedded in an application context.

While there has been widespread acknowledgment and acceptance of the *NCTM Standards*, how to actually implement change in the classroom on an everyday basis remains problematic. The *NCTM Standards* are learning goals. They are not a curriculum detailed enough to enable regular classroom mathematics teachers to implement *standards*-based lessons on an everyday basis, 180 days a year. Recognizing this problem, the [National Science Foundation](#) responded in 1989 by supporting the development of thirteen NCTM standards-based mathematics curricula materials projects, K-12, that are intended to be "full replacement" texts. At the high school level, these texts include:

- [*The Interactive Mathematics Program \(IMP\)*](#),
- *CORE-Plus Mathematics Project*,
- *Applications Reform in Secondary Education (ARISE)*,
- *Math Connections*, and
- *Systemic Initiative for Montana Mathematics and Science (SIMMS)* .

(For a synopsis of the common features of these NSF-sponsored mathematics curricula and their rationale, see Appendix A. [Common Features of an NSF Curriculum](#))

The Interactive Mathematics program (IMP)

This report details the research results on student achievement using the *Interactive Mathematics Program (IMP)* in Philadelphia public schools. *IMP* was one of the first NSF-sponsored, standards-based, full-replacement, high school mathematics curriculum projects. Like

other new curricula sponsored by the NSF, IMP is a high performance curriculum requiring instructional practices that deeply embody *NCTM's Curriculum, Teaching and Assessment Standards*. IMP consists of 20 highly contextualized thematic units built around large problems. (See Appendix B, [Detail of IMP Math Topics](#))

The IMP writing team consisted of two mathematicians from San Francisco State University, Dan Fendell and Diane Resek, and two mathematics educators from the EQUALS program at University of California at Berkeley, Sherry Fraser and Lynne Alper. The design parameters for this new curriculum were as follows:

1. It had to fully embody the content recommendations and spirit of the 1989 National Council of Teachers of Mathematics *Curriculum and Evaluation Standards*.
2. It had to be mathematically challenging for the best and the brightest students.
3. It had to be a curriculum accessible to all students.

The first IMP units were written in 1989 and were field tested in a limited number of classrooms in three high schools in the Berkeley, California area. Over the next several years, as more units were written, other pilot sites were added. Subsequent feedback from IMP teachers, including Philadelphia teachers, resulted in numerous rewrites for each unit prior to the finished product. The 9th grade level, or first year of IMP, became commercially available in August 1996. The fourth year of IMP became commercially available in August 1999. In short, it has taken ten years to fully write and field test the complete four years of IMP.

IMP in Philadelphia

In March 1992, a four and one half year contract was awarded by San Francisco State University Foundation (SFSUF) to [PATHS/PRISM: The Philadelphia Partnership for Education](#), a local education fund, for the purpose of administering the dissemination of the pre-publication pilot

version of *The Interactive Mathematics Program*. The contract was funded by a grant from the National Science Foundation to SFSUF.

IMP in Philadelphia began in the 1993-94 school year with nine teachers representing six out of thirty-five public high schools. Approximately 300 9th grade students out of a total Philadelphia 9th grade student enrollment of 15,000--roughly 2%--were enrolled in IMP the first year. (See Appendix for a time line of IMP's implementation in Philadelphia. [Philadelphia Expense Timeline](#))

Because the IMP authors retained copyright control over the dissemination of the pre-publication IMP unit booklets, the implementation standards of IMP could be set very high. All initial IMP teachers received the following:

1. ten days of training per year in each level of IMP (240 hours total),
2. on-going classroom mentoring,
3. a course load reduction of 1.5 periods,
4. a period where two IMP teachers would team-teach,
5. a classroom set of graphic calculators and LCD overhead,
6. a classroom set of manipulatives and other supplies,
7. regular citywide follow-up teacher meetings.

The above implementation standards were fairly uniform wherever IMP was being piloted in the country. The IMP authors would not grant any school permission to use the pre-publication copyrighted materials unless they agreed to these standards. In addition, in Philadelphia, four part-time co-directors facilitated the implementation of IMP. These local co-directors performed a variety of tasks which included: co-teaching IMP classes, mentoring other teachers, helping to procure materials, collecting and analyzing student achievement data, helping to prepare school

budget for the program, and generally ensuring district fidelity to the implementation model. (See Appendix under [Implementation Standards](#))

Prior to 1996, there existed neither Philadelphia nor Pennsylvania State math standards, nor any standards-based testing. There were also neither Philadelphia nor Pennsylvania accountability systems. In short, there were no external sanctions/incentives based on student achievement data. As a result, despite the many national reports released during the 1980s that documented the deficiencies in mathematics education as it was currently being taught, principals and other administrators had difficulty advocating the local need for change among their mathematics teaching staffs. In this environment, teacher recruitment for IMP relied on “heroic volunteers.” Recruiting new IMP teachers proved difficult in part because, in order to take part in the program, teachers were required to undergo extensive professional development and additional preparation. The reduction in teacher load was at best a mild incentive for teachers to participate, as this was the only way to compensate them for the considerable increase in preparation time associated with being the first to pilot a new program.

The cost of training a teacher in IMP for four years in the early years in Philadelphia was approximately \$102,000. Roughly 80% of this cost was for the 1.5 teaching period reduction in course load. Later, the 1.5 period reduction per teacher was discontinued. By the 1999-2000 year, neither the School District of Philadelphia nor any of the surrounding suburban districts were providing their teachers with a reduction in course load. Many schools adopted various models of intensive block scheduling, which provided a reduction in teaching load. Nonetheless, even without a course load reduction of 1.5 periods per teacher, the costs of implementing a new standards-based curriculum are considerably greater than merely purchasing a new textbook series. (See Appendix on [Four Year Implementation Cost](#))

Over the years, the number of Philadelphia IMP teachers has steadily grown to involve nearly 20% of all the high school staff. However, the first school to adopt IMP for its entire student body was in the Philadelphia suburbs—not the city. Strath Haven High School in the Wallingford/Swarthmore school district was the first high school to go all IMP in 1996-97. By September 2000, approximately 20 Philadelphia suburban districts were using IMP or other NSF standards-based materials in both their middle and high schools.

Questions about Impact and Attribution: Are Better Results Really Due To The New Curriculum?

It is difficult to construct and administer a single test or measure that can conclusively demonstrate superior student achievement using IMP *and* show that it was due to IMP and IMP alone. We have heard a variety of concerns regarding the impact of IMP and issues involving attribution, that is, related to cause and effect. For example, **“maybe you got better results because...”**

1. You “stacked the deck.” The IMP students were better students to start. They were brighter, or previously had more math courses in middle school, or were somehow different, which biased the results in favor of IMP.
2. You had the school’s better teachers teach the new curriculum.
3. IMP teachers graded their students more leniently. Students could get wrong answers and still pass the course because you used “alternate assessments,” which don’t test for right answers.
4. The gains in student achievement were due to increased professional development of the IMP teachers. If you give teachers who teach pre-standards curriculum the same amount of professional development you gave IMP teachers, how do you know you would not get similar results?
5. The effect is due to the use of collaborative learning groups, not the curriculum per se.
6. It could be the “Hawthorne effect.” You got better results because it was new and students and teachers were getting increased attention.
7. The IMP teachers have reductions in course load.
8. You selected only those test questions that favored the new curriculum.

9. You got the results because you tested only low achieving students. The higher achieving students don't need IMP.
10. You got the results because you only tested IMP with the "better" students [delete 'kids']. Lower ability students would not be able to do this kind of math.
11. Colleges may not accept these new programs such as IMP, rather than "algebra." And if they do, students will not be adequately prepared for a traditional college math course. Most colleges do not let students use graphing calculators so they will be at disadvantage.
12. Students who take these new curricula in high school are regularly placed in remedial math class in college because their basic skills are so low.

All of the above statements have plausible validity even if they have no basis in fact. They need to be taken seriously and addressed. In addition, different constituencies will value different indicators of success more than others. In short, the type of data collected has to be "psychologically real" for each constituent group to be accepted by that group. For example, a principal of a special admission school said he would only consider IMP successful if it could raise SAT scores by 30 points. Another principal expressed concern about attendance rates and preventing school dropouts, while another was concerned about scores on the [Pennsylvania System of Student Assessment \(PSSA\)](#).

In response to these methodological issues and the different values placed on different measures of success, our strategy has been to collect student achievement data from a variety of sources using a variety of indicators. At the same time, we have attempted to control or compensate for as many biasing factors as possible. Our goal was and is to determine whether there is a convergence of the data in any one direction that could be used to answer the bottom line question of the costs and benefits of using a standards-based curriculum, such as IMP .

We began collecting data during the first year of IMP's implementation in Philadelphia in May 1994. Since then, we have used a variety of measures, which are listed in the Appendix under [Student Achievement Measures](#). There are several reasons to believe the superior results of NCTM standards-based curriculum and student centered instructional methods as shown with IMP in Philadelphia are actually *understated*.

1. All of the results documented herein were from pre-publication draft versions of IMP.
2. All of the IMP teachers were teaching the curriculum for the first time.
3. None of Philadelphia's most able students were permitted to enroll in IMP at each school.
4. None of the 9th grade students had previously been exposed to standards-based k-8 materials.
5. Most of the results were based on students using IMP for only 2 and 2/3 years.
6. IMP teachers did not operate in a "reform teaching culture," but were often isolated in their departments. The same was often true for IMP students.
7. Parents were unfamiliar with a standards-based curriculum and some did not know how to help their children.

Based on our experience, we hypothesize that significantly greater student achievement results would accrue if high school students who are taught using the IMP materials:

- a) were first taught using a k-8 standards-based mathematics curriculum,
- b) use the finished published version of the IMP text,
- c) are taught by teachers who had several years of experience teaching the same level of IMP,
- d) are enrolled in a school that had adopted a standards-based mathematics program for all its students.

Implementation Conclusions

The \$21,000 per teacher cost to implement IMP, without the reduction in teaching load, is fairly accurate based on our experience. To properly learn how to teach IMP, teachers need 40 days of training (either on school time or compensated time), in-classroom mentoring, a classroom set of graphics calculators, a closet of classroom materials, overhead projector and LCD panel and, of course, IMP textbooks. These are hard and fast requirements. Assuming the average salary and benefits of teachers are at least \$50,000 per year or \$200,000 over four years, this \$21,000 represents a 10% investment over and above normal personnel costs.

The reduction in teaching load combined with team teaching is the ideal model. But much of the benefits of that model can be achieved through a semester block schedule, which has a built-in reduction in student load, and rostering a common prep period for teachers. All of a mathematics' department teaching staff should be trained in the IMP curriculum for all four years to prevent loss of program capacity in the event of teacher transfers, retirements, extended absences and deaths. In addition, the more mobile teachers are within a large district, or within a region, the more advantageous it is to train all of a district's or region's math teachers in either IMP or a comparable NSF sponsored standards-based curricula.

There are other considerations in implementing systemic change, such as preparing teachers for change, articulating with higher education, systematizing data collection and program evaluation, updating hiring practices, establishing different teacher evaluation protocols, training and mentoring new teachers, and on-going administrative education.

Systemic change is more likely to occur in an educational environment that includes mandated mathematics curriculum standards, revamped assessment including performance assessments, and a strong school and district accountability system based on student achievement.

However, we have found that without strong top-level administrative support and direction, systemic change will not happen, despite these external inducements and sanctions.

Change must reach into the classroom involving every member of the mathematics teaching staff.

Utilization of research-based curriculum and instructional methods should not be an option teachers could reject. Conversely, merely purchasing these textbooks is insufficient and even counter-productive.

What follows is a summary of the research results of the Philadelphia experience implementing the *Interactive Mathematics Program*.

Philadelphia IMP Research Summary

Early Philadelphia Data: 1994, 1995, and 1996.

1. Student Attitudinal Surveys, 1994: As a new program, which was completely voluntary both for the IMP teachers and students, the first question we wanted to know was whether the 9th grade students who chose to take IMP would want to continue with the program. If no students wanted to continue, we could not run the program. To gauge student interest, we developed a short questionnaire on content and pedagogy and gave it to students at the end of their first year, in May 1994. We wanted to know whether there were any large differences between different types of schools. We had started with three high achieving “special admission schools,” one vo-tech, and two neighborhood comprehensive schools.

The result: students overwhelming preferred IMP to algebra classes and indicated they would continue with the course. (See the [Ninth Grade Student Attitudes](#) report in the Appendix.)

2. Passing Rate and Attendance Data: Principals in Philadelphia’s twenty-two comprehensive high schools face the problems of low student passing rates and a year-by-year loss of students who drop out of school. In most of these high schools, the graduating class is less than half of what it was in ninth grade. As IMP expanded into these comprehensive high schools, the first question was whether IMP students had higher passing and attendance rates than did students using pre-standards materials.

The results: By and large, IMP students had consistently higher passing and attendance rates in all of the comprehensive schools, not only in math, but English, social studies and science compared with their peers in the same schools. (See [Passing Rate Comparisons 1995, 1996, 1997](#))

Methodological note: To guard against the claim that IMP teachers were liberal graders, thereby inflating the IMP passing grades, we also looked at other subjects and found higher passing rates in those subjects as well (English, science and social studies). We attribute these results to the writing and problem solving emphasis in IMP, which seemed to have a secondary effect on other subjects.

However, we did not have 8th grade test scores to compare the two groups and control for cohort effects. Thus, another possible interpretation for the higher IMP results was that we had “better” kids to start. This was a valid concern, which we would systematically address in later studies. But for now, we were encouraged by these early results because none of Philadelphia’s honors’ students were allowed to take IMP. Also, IMP students within a school were usually randomly assigned or they self-selected.

3. PSAT comparisons: IMP began in three special admission schools: Central High School, the second oldest high school in America and formerly an all boys school, Philadelphia High School for Girls (still all girls), and Carver High School of Engineering & Science. These schools draw students from the entire 8th grade student population in Philadelphia and, as a consequence, attract the top scoring students. For the principals in these schools, the main issues are SAT and Advanced Placement scores rather than passing rates. As one principal remarked, “I’ll consider IMP successful if it can raise SAT (Scholastic Aptitude Test) scores 30 points by the junior year.”

The SATs were originally designed in the 1930s to be objective criteria upon which to base admissions to college. As such, its purpose was to measure the *aptitude* of the student that was separate and apart from any previous instructional influences. It was expressly designed *not* to measure achievement per se. The Math and verbal sections of the SAT, therefore, function more like an intelligence test, rather than as indicators of achievement in mathematics or literature and

composition. Certainly the design of the Interactive Mathematics Program was geared to prepare students for deeper mathematical understanding and higher achievement, not for higher SAT scores per se. Nonetheless, the misuse of the SATs to compare the effectiveness of schools and curricular programs is a political reality. We, therefore, had to collect comparative data.

In October 1994, we began to collect PSAT data from our first Philadelphia IMP cohort of students in Central, Girls and Carver. These students were sophomores who had completed one year of IMP. They had not been the highest achieving students in 8th grade. Those students were not given the option of enrolling in IMP.

Results: IMP sophomore students outperformed traditionally taught students in all three special admission schools in *both* the math and verbal sections of the PSAT. (See [PSAT Score Comparisons 1994, 1995, 1996](#) in the Appendix)

In October 1995, the original cohort of IMP students was again analyzed for their 11th grade PSAT scores. Strawberry Mansion, an inner city comprehensive high school, was also included.

Results: IMP juniors continued to outperform traditionally taught students in the special admission schools and one comprehensive school in *both* the math and verbal sections of the PSAT.

Methodological Notes: This analysis does not control for cohort effects, that is, 8th grade achievement levels, nor for the effects on PSAT students who take more math courses, particularly algebra in 8th grade. However, for reasons previously outlined, these biases would tend to favor the traditional students, many of whom, unlike the IMP students, had taken algebra in the 8th grade and went on to take geometry in 9th grade and algebra II in 11th grade. These students were essentially one year ahead of the IMP students.

Later Philadelphia Data: 1996, 1997, 1998.

4. Stanford Achievement Test – 9th Edition (SAT-9)

In August 1994, David Hornbeck began his six-year tenure as the Superintendent of the School District of Philadelphia. With backing from the school board, he initiated a 10-point “Children Achieving” agenda. A key part of his agenda was greater accountability for student performance. A criteria referenced version of the Stanford Achievement Test- 9th edition (SAT-9) was developed to be the centerpiece a new system of accountability. Three subjects were tested: reading, math, and science. The first “baseline” SAT-9 test was administered in April, 1996.

The results of the SAT-9 test for each subject area were separately reported to the district on a 0-99 scale as normal curve equivalents. The district then converted these scores into a seven level rubric. A “0” was untested and a “Level 6” was considered “Advanced.” Then the School District of Philadelphia assigned a numerical weight to each level: *Advanced* = 1.2, *Proficient* = 1.0, *Basic* = .8, *Below Basic III* = .6, *Below Basic II* = .4, *Below Basic I* = .2 and *Non Tested* = 0. These weights were then multiplied by the percentage of a school’s students who scored in each of these seven levels for a particular subject. The total point values for all levels were then added together. This result was the performance index score for that subject. (Equivalently, the students’ scores reported on the 0 through 6 scale are multiplied by 20 and then averaged to obtain the performance index.) For example, if 100% of the students took the test and 100% scored at the advanced level, Level 6, the performance index would be 6 x 20 or 120. On the other hand, if 100% of the students did not take the test the performance index would be 100% x 0 or 0. In 1996, the actual math performance index score for all Philadelphia public high school was **27.2**. See the Appendix for the [Stanford 9 Test and Performance Index](#).

The same calculation was also performed for reading and science. Two additional indices were also developed: one for student and staff attendance, and another for “student persistence.”

These five index scores were added together and averaged. The result was a school's "Performance Index." The goal was to achieve a Performance Index of **95** in twelve years. A team of central office personnel reviewed principals whose schools that did not show progress toward meeting this twelve-year goal. Within a few years, the SAT-9 became a very high stakes test for building level administrators. However, initially, many teachers and students did not invest much time in the administration of the test. Overall, in 1996, half of the students did not take the test. The exceptions were the special admission schools, such as Central, Girls and Carver, whose students tend to take all tests more seriously. In the 1996 SAT-9 testing, these schools had a relatively low number of untested students.

Using the March 1996 SAT-9 test results for Central and Girls High, we did a matched sample comparing IMP juniors to a similar group of pre-standards, traditionally taught students from these same schools. We matched by gender, whether they came from public or private elementary schools, and their 8th grade national percentiles in the math and verbal sections of the Comprehensive Test of Basic Skills (CTBS). Our sample sizes were 83 IMP students at Central and 55 IMP students at Girls matched against 83 and 55 traditionally taught students respectively. We analyzed the raw scores on each item of the math, reading and science sections for the SAT-9.

Results: At Central High school, IMP students outperformed traditionally taught students in 21 out of 28 reported SAT-9 multiple-choice categories. Of these Probability and Functions gain scores were statistically significant. There were 3 categories that were ties. There were 4 categories where the traditionally taught scores were higher than IMP. Of these, none was statistically significant.

Results: At Girl High school, IMP students outperformed traditionally taught students in 12 out the 17 math-related sub-scores, tied on two and scored slightly lower on three. IMP students did better on all the cumulative scores and open-ended assessments. Moreover, only 29% of the IMP students' scores "below basic" compared to 43.6% of the traditionally taught students.

(See Appendix for [Central and Girls' SAT 9 Matched Sample Analysis](#))

In the Spring 1997 administration of the SAT-9 test, the percentage of untested students dropped from 49.9% in 1996 to 30.2% in 1997. Accordingly, the math performance index score rose from 27.2 in 1996 to 36.8 in 1997. In short, many principals could realize their targeted performance gains for the first two-year reporting cycle merely by getting more students to be tested.

Using the Spring 1997 data, we performed another study comparing IMP students to traditionally taught students in three special admissions schools and from several comprehensive high schools. We obtained students' 8th grade CTBS scores to control for cohort effects. In the appendix [1997 SAT 9 Matched Sample Analysis](#) we present the school-by-school results for all of the students who were tested. Then we did a more carefully matched sample study across the magnet or special admission schools comparing 96 IMP and 96 traditionally taught students, whose average CTBS scores were nearly equal. We also compared 167 IMP and 167 traditionally taught students in the four comprehensive high schools whose CTBS scores were nearly equal. (In both cases, IMP students had slightly lower 8th grade CTBS scores.) We reported the scores using the school district's 0-6 level rubric.

Results: IMP students scored *higher* than traditionally taught students whether they were in special admission high schools or comprehensive high schools. This translates into a math performance score gain of 5.4 points for the special admission schools and a 5.2 performance point gain for the comprehensive IMP students as matched against a comparable group of their peers.

In the Spring 1998 administration the SAT-9 test, principals continued to make a concerted effort to get more of their students tested and to encourage them to take the test seriously. The percentage of untested students dropped to 26.2%, and the performance index score for math rose to 40.0---still 55 points under the 95-point target. With the third administration of the SAT-9, two realities emerged: 1) the use of this test as part of a accountability system was unlikely to go away, and 2) there were ceiling limits on how much gain could be achieved through improved testing

conditions and incentives for students. As a result, principals had increased incentives to examine changing their curriculum and instructional methods in the three major subjects tested in their schools. This gave rise to practical questions about the extent to which IMP was appropriate for different sub-groups of students: ESOL, Special Education, Advanced Coursework, and Mentally Gifted.

By the Spring 1998, there were 10 Philadelphia high schools that had juniors who took the SAT-9 test. We analyzed SAT-9 results for each school and then performed a series of aggregate analyses, which are presented in the appendix as [1998 SAT 9 Charts 1A to 4C](#). In these 10 schools, we had a total IMP junior student test population of 407 students and a non-IMP junior test population in those same schools of 2200 students. We then compared these two groups across nine sub-scores for the SAT-9 in reading, math and science. We used the 0-6 rubric scale.

The results: IMP students outperformed non-IMP students in every sub-score, including reading and science. The math Performance Index for IMP students was 11.8² points higher than for non-IMP students. (See Chart 1-A)

What if the IMP students were better students in 8th grade? We then looked at only those students who had 8th grade CTBS scores, which included 232 IMP and 1,018 non-IMP students. They were roughly equal in their 8th grade CTBS scores: 778 versus 773. (See Chart 1-B.)

The results: IMP students outperformed non-IMP students in every sub-score, including reading and science. The math Performance Index for IMP students was 11.4 points higher than for non-IMP students. (See Chart 1-C).

We then separated out all the students who had taken advanced course work in math, all the special education students, mentally gifted and students for whom English was not the primary language. We were left with 355 IMP students and 1,672 non-IMP students.

The results: IMP students outperformed non-IMP students in every sub-score, including reading and science. The math Performance Index for IMP students was 16.8 points higher than for non-IMP students. (See Chart 2-A.)

² This is calculated by taking the difference in the “Math Composite scores” (3.09 vs. 2.5) and multiplying it by 20.

From this group, we further sorted out all those students who had 8th grade CTBS scores. This reduced the number to 194 IMP students versus 761 non-IMP students. The IMP students were somewhat higher in 8th grade CTBS scores, 771 versus 758. (See Chart 2-B.) We then analyzed this sub-group's SAT-9 scores.

The results: IMP students outperformed non-IMP students in every sub-score, including reading and science. The math Performance Index for IMP students was 15.6 points higher than for non-IMP students. (See Chart 2-C.)

Next we took the original population (Chart 1-A) and analyzed only those students from Central, Girls and Carver high schools. We excluded the mentally gifted, ESOL, and those who had advanced coursework. This left us with a total of 112 IMP students and 430 non-IMP students across all three of these special admission schools.

The results: IMP students outperformed non-IMP students in every sub-score, including reading and science. The math Performance Index for IMP students was 7.8 points higher than for non-IMP students. (See Chart 3-A.)³

From this special admission school population we then took only those students who had 8th grade CTBS scores. This reduced the sample to 56 IMP students and 178 non-IMP students who scored 812 and 804 respectively on their CTBS tests. (See Chart 3-B.)

The results: IMP students outperformed non-IMP students in every sub-score, including reading and science. The math Performance Index for IMP students was 7.8 points higher than for non-IMP students. (See Chart 3-A.)⁴

We did a similar analysis, as displayed in Charts 3-A, 3-B, and 3-C, for the comprehensive school students. This population, regardless of whether they were IMP or non-IMP students, scored substantially below their special admission school peers. We aggregated all the comprehensive

³ The aggregate IMP Reading Opened subscore is lower than the non-IMP does. However, in each of the three schools taken separately, IMP students had, in fact, higher opened ended reading subscores.

⁴Ibid.

students, but excluded the ESOL, special education and advanced students. This left us with a sample of 243 IMP students and 1,212 non-IMP students.

The results: IMP students outperformed non-IMP students in every sub-score, including reading and science. The math Performance Index for IMP students was 18.6 points higher than for non-IMP students. (See Chart 4-A.)

We further narrowed the sample by selecting only those students who had 8th grade CTBS scores. This gave us a reduced sample of 138 IMP students vs. 568 non-IMP students. Their CTBS scores were 755 and 742 respectively. (See Chart 4-B.)

The results: IMP students outperformed non-IMP students in every sub-score, including reading and science. The math Performance Index for IMP students was 17.4 points higher than for non-IMP students. (See Chart 4-C.)

ANCOVA study of Girls High data: One methodological problem of using matched samples or matched pairs to compare IMP to non-IMP students is the potential criticism that in choosing two equally good non-IMP “matches” to a particular IMP student, the reader does not know if the researcher biased the results by selecting the weaker non-IMP match.

The advantage of doing an ANCOVA is that we can use all of the data even if the IMP and non-IMP groups are not perfectly matched in either their 8th grade test performance or the respective sample sizes. In an ANCOVA, we used essentially two regression lines. The first plotted each SAT-9 score (using normal curve equivalents, which ranged from 0-99) as a function of each student's 8th grade CTBS score. This gave us a “line of best fit” for the non-IMP students. If there were no predicative power in knowing a student’s CTBS score, the line would be flat, i.e. horizontal. How closely the points cluster about the line determines the strength of the linear relationship. We also did a line of best fit for the IMP students at Girls High.

If the IMP students’ SAT-9 scores were really better than the non-IMP students (after controlling for 8th grade scores) then the IMP line of best fit would lie above the non-IMP line. Moreover, by comparing the “slopes” of each line one could tell if there were any trends between

higher and lower initially achieving students. For example, if the IMP line were above the non-IMP line and if the distance between the two lines progressively increased as the 8th grade scores increased, it would indicate that the higher scoring students benefited the most from taking IMP.

Results: Girl's IMP 11th grade students significantly outperformed their non-IMP counterparts on the SAT-9. Furthermore, there was a trend such that the higher achieving IMP students tended to perform did even better. (See [1998 Girls' Ancova Study](#))

Methodological Note: Because the test items of the CTBS and the SAT-9 are not aligned, our preference would be to use students' 8th grade Stanford-9 scores as baseline data and then compare their 11th grade Stanford-9 scores. The earliest we could have done this analysis was with the 1995-96 cohort of 8th graders who would have taken the 8th grade version of the SAT-9 in April 1996 (its first year in Philadelphia) and the 11th grade version three years later in April 1999. As of the date of this report, we are currently in the process of conducting such an analysis, and we hope to present the results in PART II of this research report.

6. Ivy League University Math Exit Test (an undergraduate degree requirement). For college bound students, the questions we have heard, beyond college admissions issues, center around how well students will be prepared for higher-education coursework. Will high school students who take programs like IMP suffer once they get into college because of IMP? Will they take disproportionately more remedial math courses? Will their choices for technical majors be restricted? How will they fare in a traditional college math course? What if they are not permitted to use graphics calculators? Will their GPAs suffer as a result?

While these questions are legitimate concerns of parents and educators, they are asked against an existing backdrop of statistics. We know there has already existed to varying degrees the need for math remediation in four-year and two-year colleges. In part this is due to more students

aspiring to college and to higher requirements for math achievement. Prior to programs such as IMP, there has been a chronic shortage of American college students pursuing technical majors, as well as high college drop out rates. Thus, a key research problem is correct attribution. For example, one Philadelphia math department head criticized IMP for supposedly producing students who needed remedial math at a nationally known Philadelphia university where he taught as an adjunct calculus instructor. But when he was asked to distribute a survey to his 200 incoming freshman, it turned out that none of these students had been IMP students, nor had any participated in any other NSF curriculum math program.

Undoubtedly there will be IMP high school students who later need to take remedial college math. Some students are weak regardless of their math program in high school or have chronically poor study habits. Some IMP students will shy away from technical majors, will not fare well in traditional college courses, will have low GPA scores and may even drop out. In short, there will be students who fare no better than the status quo. The real question is whether IMP students *taken as a group* will have a *greater* tendency to do better or worse in college than their non-IMP peers, all things being equal.

To get a quantitative handle on the issue of college preparedness, we invited Norman Webb, a nationally renowned educational researcher from the University of Wisconsin, to conduct a study. Dr. Webb selected ten questions from a quantitative reasoning test given to undergraduate college students at an ivy-league university. Among the topics covered by the questions were statistical reasoning and interpretations of graphs. We then administered these to 150 students from Central High School in June 1996 who had just completed their junior year. Of these 150 students, 91 had been enrolled in IMP for three years. (This was IMP's first student cohort.) 59 students were enrolled in a traditional Algebra II class. Each group's 8th grade CTBS scores were compared and found to be nearly equal.

Results: The IMP students got a little more than 50% of the questions correct whereas the non-IMP students got less than 25% of the questions correct. This difference in performance was statistically significant to the $p < .0001$ level.

Dr. Webb and his colleague Maritza Dowling performed a more in-depth precise matched-group analysis with an accompanying report, which is included in the appendix. A copy of the test used is also included in the appendix. [College Quantitative Reasoning Study](#)

Conclusions

The Interactive Mathematics Program curriculum materials, however well crafted, will not by themselves yield increased student achievement. It is not a teacher proof curriculum, and so this program is not about buying new textbooks. It is rather a profoundly intense, multi-year professional development program exploring rich, varied and meaningful mathematical content and subtle yet profound pedagogical techniques to promote deeper student understanding. When implemented systemically, IMP can significantly enhance student understanding of and competence in mathematics. IMP also has been demonstrated to have collateral benefit in reading and science as well.

APPENDICES

- The Philadelphia Experience - Timeline
- Our Implementation Standards
- Four Year Implementation Costs of a NSF High School Exemplary Curriculum
- Student Achievement Measures
- Early Philadelphia Data 1994, 1995, 1996
 - Student Attitudinal Surveys
 - Passing Rate and Attendance Data
 - PSAT comparisons
- Later Philadelphia Data: 1996, 1997, 1998
 - Stanford Achievement Test – 9th Edition (SAT-9)
 - Ivy League University Math Test
- A. Common Features of an NSF-sponsored Curriculum
- B. Description of the Interactive Mathematics Program
- C. Elements of Systemic Change
- D. The Greater Philadelphia Secondary Mathematics Project
- References Cited

PART II: Pending Research: 1999, 2000, 2001, 2002

- Strath Haven Studies: ERBs, SATs, PSATs, CORE-Plus Algebra Test, NAEP
- New Standard Reference Exam
- New York Math Regents Exam (New Year City Research Study)
- Philadelphia Community College Analysis
- New Jersey GEPA + HEPA state tests
- Pennsylvania System of Student Assessment (PSSA)

THE PHILADELPHIA EXPERIENCE - TIMELINE

1. **March 1993:** Received sub-award from National IMP in San Francisco to establish the *Philadelphia Interactive Mathematics Program (IMP) Regional Dissemination Site*.
2. **Summer, 1993:** Training begins with 9 teachers and 4 co-directors. Six Philadelphia schools comprise the first IMP cohort: Central, Girls, Carver, Dobbins, Gratz and Strawberry Mansion High Schools.
3. **August, 1993:** Philadelphia Superintendent Constance Clayton retires after 10 years as Philadelphia Superintendent. First attempt to write a Philadelphia USI.
4. **August 1994:** Philadelphia IMP expands to 9 schools. David Hornbeck hired as Philadelphia's new superintendent. Philadelphia resubmits USI proposal as part of Mr. Hornbeck's "*Children Achieving Agenda*." *Philadelphia Regional IMP Site* moves to LaSalle University. Philadelphia begins drafting its new Curriculum Standards.
5. **August, 1995:** NSF awards *USI* grant to Philadelphia. PHUSI develops partnership with *Philadelphia Regional IMP Site* at LaSalle University and supports IMP expansion. IMP now in 12 Philadelphia high schools.
6. **Spring, 1996:** Philadelphia first uses the Stanford Achievement Test- 9th edition (SAT-9) as criterion referenced test in math, English and science, to establish baseline scores in grades 5, 8 and 11. New school accountability system adopted, which is 60% based on SAT-9 scores. Pennsylvania begins new state assessment system for students in grades 5, 8 and 11.
7. **Summer, 1996:** Strath Haven High School, in prestigious Wallingford/Swarthmore School District, becomes the first suburban high school to adopt IMP in the Philadelphia area and first to commit to go all IMP. SHHS IMP is not grant supported, but contracts for in-service. Philadelphia IMP begins training nine New York City teachers in IMP on contract with the *New York City USI*.

8. **August, 1997:** Contract between *National IMP* and *Philadelphia IMP* ends. *Philadelphia USI* supports entire *Philadelphia IMP* operation at La Salle University pending its LSC proposal to NSF. IMP expands to 20 Philadelphia high schools.
9. **June, 1998:** NSF awards LSC grant to La Salle University. Former Philadelphia IMP directors continue as LSC directors for *The Greater Philadelphia Secondary Mathematics Project (GPSMP)*. *GPSMP* expands IMP in the suburbs, expands to include *CORE-Plus*, and expands to include middle schools using *Math in Context (MiC)* and *Connected Math Program (CPM)*.
10. **January, 1999.** Pennsylvania adopts mandatory content standards in math and reading.
11. **July, 1999:** The *GPSMP* expands summer training to three more suburban school districts that begin a four-year process of adopting NSF curricula whole school for all their secondary math staff and students. (Bethlehem, Haddon Township and Pennsauken School Districts). NSF awards two supplemental grants to *GPSMP*: “*Systemic Elementary Mathematics Teacher Tutoring Initiative*” (*SEMTTI*) and the *Strath Haven Research Study*.
12. **February, 2000.** The *GPSMP* expands to involve 20 school districts, approximately 90 middle schools and high school schools, and 800 teachers in a multi-year systemic change process. New York City IMP schools number over 20. The *Philadelphia USI* continues to contract with *GPSMP* to provide IMP mentors to its comprehensive high schools. Other reciprocal agreements are made between *GPSMP* and *PHUSI*. The *GPSMP* continues its fourth year of providing IMP training to New York City teachers in IMP. The original NYC IMP teacher cohort is now training new NYC IMP teachers. The *GPSMP* staff now numbers over 90 people (5 full time and 89 part time).
13. **November 2000.** Strath Haven High School, the first all IMP high school in the Philadelphia area, scores tops on Pennsylvania System of School Assessment (PSSA) for a comprehensive high school and number 2 overall out of over 600 high schools statewide.
14. **March, 2001,** *GPSMP* staff helps the Bronx beginning training 500 high school teachers in either IMP or Math Connections.

**IMPLEMENTATION STANDARDS FOR
NATIONAL SCIENCE FOUNDATION-SPONSORED
EXEMPLARY SECONDARY MATHEMATICS CURRICULA**

1. PROFESSIONAL DEVELOPMENT: 240 HOURS OF TRAINING
(60 HOURS PER YEAR X 4 YEARS)
 - a. New Mathematical Content e.g., statistics, probability
 - b. Novel Problems in Unfamiliar Contexts
 - c. Inquiry-Based, “Active Learning” Instruction
 - d. Graphing Calculators Used Extensively
 - e. Alternative Assessments
2. CLASSROOM MENTORING VISITS OVER 4 YEARS: 50 HOURS PER TEACHER
 - a. Demonstration Lessons
 - b. Pre and Post-class conferencing
 - c. Quality Assurance
3. TIME FOR TEACHER-TO-TEACHER COLLABORATION & NETWORKING
 - a. Joint-Planning Period During or After School
 - b. National and Local List Serves
4. TEACHER COMPENSATION/INCENTIVES
 - a. 10 in-service days per year at 6 hours per day.
 - b. Extra planning to learn and teach new material
 - c. Student-teacher assignments

5. REGULAR TECHNICAL ASSISTANCE TO SCHOOL ADMINISTRATORS
 - a. Budgeting;
 - b. Teacher, student recruitment issues;
 - c. Appropriate teacher and student roster configurations;
 - d. Transfer, absentee and retention issues;
 - e. Classroom materials, books, calculators requisitions;
 - f. Student testing and college preparation issues;
 - g. National information updates
 - h. Public relations, informational presentations
6. APPROPRIATE TEACHING ASSIGNMENTS & STAFF STABILITY
7. DISTRICT POLICY AND RESOURCE ALIGNMENT
 - a. New Hires and Transfers
 - b. Teacher evaluation criteria
 - c. Testing formats and accountability policy
 - d. School-based and district resource alignment
8. STUDENT ACHIEVEMENT PROGRAM EVALUATION FOR EACH SCHOOL SITE
9. MECHANISMS FOR RECRUITING NEW TEACHERS AND SCHOOLS
10. HIGHER EDUCATION ARTICULATION
 - a. Admissions
 - b. New Teacher Preparation
 - c. Accessing Collegiate Resources

F. Joseph Merlino, Project Director
Greater Philadelphia Secondary Mathematics Project, June 1999

**Four Year Implementation Costs of a NSF
High School Exemplary Curriculum per Teacher
YEAR 2000**

Expense Category	Calculations	Costs
New Textbooks	\$42/book x 35 students x 5 classes	\$7,350
Graphing Calculators	\$80 x 35 students (one classroom set)	\$2,800
Overhead Projector	One	\$240
Overhead Graphing Calculator	One	\$250
Classroom Materials	One kit per teacher	\$300
Teacher in-service stipends	\$25/hour x 60 hours/year x 4 years	\$6,000
Sub-Total School Share		\$16,940

Professional Development	\$400/yr per teacher prorated x 4 years	\$1,600
Classroom Mentoring	\$800 1 st year + \$400 in years 2, 3, and 4	\$2,000
Technical assistance to schools	12 person days x 300 per day prorated	\$900
Sub-Total NSF/LSC Share		\$4,500

TOTAL COSTS TO IMPLEMENT FOR ONE TEACHERS	\$21,440
--	-----------------

Original Model		
Teacher release period for team teaching	1.5 periods x \$13,500/per x 4 years	\$81,000
Revised total using original IMP model		\$102,440

Average salary & benefits per teacher	\$64,000/year x 4 years	\$256,000
---------------------------------------	-------------------------	------------------

STUDENT ACHIEVEMENT MEASURES WE HAVE USED OR ARE IN THE PROCESS OF USING

Classroom Focused

- Passing Rate Data (all four quarters) (Philadelphia)
- Grade Point Averages (Central HS, pending Palisades SD)
- Attendance Rates (Philadelphia)
- Attitudinal Surveys of Students (Philadelphia + Furness HS)

Criterion Referenced, School Achievement Focused

- Stanford Achievement Test--9th Edition (SAT-9) Philadelphia
- New Standards Reference Exam (pending, Bethlehem, Palisades SD)
- New York Regent's Math Exam (pending Sequential and the "Form A")
- Core-Plus Algebra Test (pending, Strath Haven Research Study)
- NAEP statistics questions (pending, Strath Haven Research Study)
- New Jersey's Performance Assessments, 8th and 11th grade

Norm-Referenced, School Comparison Focused

- Pennsylvania System of School Assessment (PSSA) (pending)
- Education Record Bureau (ERBs) + SATs (pending Strath Haven HS)

College Admission Focused

- Scholastic Aptitude Test scores (PSATs and SATs)
- Ivy League University Math Exit Test
- Philadelphia Community College Admissions Test (pending)

Today's mathematics curriculum must prepare students for their future roles in society. The secondary courses are based on principles that are consistent with those that under-pin the elementary program, facilitating the transition from elementary school. Operations that were an essential part of a procedures-focused curriculum for decades can now be accomplished quickly and effectively using technology, so that students can now solve problems that were previously too time-consuming to attempt, and can focus on underlying concepts. In an effective mathematics program, students learn in the presence of technology. Technology should influence the mathematics content taught and how it is taught. NSF Director's statement. The Third International Mathematics and Science Study (TIMSS) was designed explicitly to enable educators and policy makers to compare achievement in science and mathematics of students in the United States with those in other countries at three levels of education, grades 4, 8, and the final year of secondary school (grade 12 in the U.S.). With this publication of the results of Improving achievement in mathematics and science subjects, whether in basic skills A1.1: Nations' Definitions Of Eligible Students And Whether Met Sampling Standards: Mathematics And Science General Knowledge Assessments . . . 86. Beginning in 1990, the National Science Foundation (NSF) funded several mathematics curriculum development projects to respond to the need for curriculum materials to be more closely aligned with the NCTM's Standards. Curriculum development teams consisted of mathematics educators, mathematicians, and classroom teachers. The curriculum materials were extensively field-tested and revised on the basis of feedback from teachers and evidence of student learning (Senk & Thompson, 2002). Information about the impact of Standards-based mathematics curriculum materials on student learning is clearly needed.