the
HIGH SCHOOL REFORM CONFERENCE

series

Using Rigorous Evidence to Improve Policy and Practice

The High School Reform Conference Series

Using Rigorous Evidence
to Improve Policy and Practice
Colloquium Report

January 22-23, 2004
New Orleans, Louisiana

Convened by

MDRC

with

The National High School Alliance
The Council of the Great City Schools

mdrc
BUILDING KNOWLEDGE
TO IMPROVE SOCIAL POLICY

For information about MDRC and copies of our publications, see our Web site: www.mdrc.org.

Contents

Preface

The state of the American high school in urban communities is a matter of great concern. Families, educators, and policymakers continue to be discouraged by high dropout rates, low academic achievement, and the failure of many high school youth to meet graduation requirements and reach the level of educational attainment necessary to enter college. The national agenda, as embodied in the 2001 No Child Left Behind Act, explicitly draws attention to low-performing schools, many of which are at the high school level, and calls for a renewed focus on high school improvement. There is now greater urgency than ever before to build scientifically based evidence that shows what works to improve student retention, graduation rates, and academic achievement. The goal is for such evidence to be used to inform policy and improve practice in the field.

Given this pressing need, and with the support and encouragement of the Bill & Melinda Gates Foundation and the William and Flora Hewlett Foundation, MDRC was delighted to have the opportunity to convene a day-and-a-half colloquium, "High School Reform: Using Rigorous Evidence to Improve Policy and Practice." This meeting brought together a wide-ranging group of leading researchers, policymakers, school administrators, and foundation officers, who spent the time grappling with the challenges of utilizing research-based evidence to inform practice and policy in urban high schools. The goal of the colloquium was not to focus on descriptions of the problems but, rather, to delve into the existing research base to discuss what has been learned so far and where the current knowledge should lead and to begin shaping an agenda for forward movement and further discussion. The commentary that follows synthesizes the major themes and key messages from the colloquium, which featured 23 speakers and panelists and was attended by 75 researchers, policymakers, and practitioners. The papers that are included, which provoked lively discussion and new ideas during colloquium sessions, are meant to provide the opportunity for further thought about the very important questions they raise. MDRC views this event as the first in a series of conferences that will deepen the knowledge and understanding of what works to improve high schools and how the evidence can contribute to better policy and practice. The aim is to build a learning community of scholars, school administrators, and policy experts dedicated to this cause.

Many people contributed to the success of this first colloquium. The conference planners gratefully extend their appreciation to the Bill & Melinda Gates Foundation and the William and Flora Hewlett Foundation for their support, and to the National High School Alliance and the Council of the Great City Schools for partnering with us to bring this concept to fruition. We also appreciate the input and guidance of our advisors: Nancy Adelman and Barbara Means, SRI International; Steve Fleischman and Rebecca Herman, American Institutes for Research; Naomi Housman, National High School Alliance at the Institute for Educational Leadership;

Colloquium Summary

Introduction

The enactment of the 2001 No Child Left Behind Act is transforming the education landscape in numerous ways. The focus on standards and accountability measures is putting pressure on schools to improve academic achievement levels. There is increased emphasis on testing and assessment, with clear consequences for schools that don't meet standards. And there is also a push for the use of scientifically based methods to ratchet up the rigor of research. Given these shifts and the issues they raise, MDRC has designed a series of conferences that aims to bring policymakers, researchers, and practitioners together to explore the challenges and benefits of using rigorous evidence to improve low-performing high schools.

The first conference in the series took place in New Orleans, Louisiana, on January 22 and 23, 2004. Sponsored by the Bill & Melinda Gates Foundation and the William and Flora Hewlett Foundation, led by MDRC, and co-convened by the National High School Alliance and the Council of the Great City Schools, the conference was structured to have the intimate feel of a colloquium. The focus was on creating dialogues about the state of research, the current general knowledge about improving high schools, the key challenges involved in researching reform, and promising strategies designed to address the challenges. Two future conferences will involve a broader group of individuals from the field, with the goal of better connecting research findings to policy and practice. But this first gathering was intended to define and discuss the research-related issues, questions, and concerns. It consisted of three sessions, which addressed three major themes:

- What has been learned so far about the effectiveness of high school reform approaches?

- How can the field of education maximize the role research can play in informing practice and policy?

- What major studies are under way to evaluate various high school reform programs and what are the emerging lessons?

Within those themes, several other important ideas emerged throughout the course of the meeting. This document aims to describe the conference proceedings; sum up the key ideas that will help shape the thinking of policymakers, researchers, and practitioners as reform efforts continue to develop and change; and set the stage for the next event.

Opening Remarks

Robert Ivry, MDRC's Senior Vice President for Development and External Affairs, opened the conference by outlining its goals. Following his remarks, four introductory speakers took the podium and set the context for the audience. **David Ferrero,** Director of Evaluation and Policy Research for education programs at the Bill & Melinda Gates Foundation, laid the basis for the colloquium's discussions by describing the foundation's reasons for supporting the series of research conferences. The largest motivating force, he said, is the need for each generation of educators to meet the challenge of reform anew. **Naomi Housman,** Coordinator for the National High School Alliance, talked about the field's hunger for information about forthcoming research, describing a new initiative from the Institute for Education Leadership, which will result in a Web site that catalogues research in progress. **Shirley Schwartz,** Director of Special Projects at the Council of the Great City Schools, discussed her organization's projects and its focus on programs designed to recruit and prepare a diverse and highly qualified teacher workforce. And **Phoebe Cottingham,** Commissioner of the National Center for Education Evaluation and Regional Assistance at the U.S. Department of Education's Institute of Education Sciences, explained the importance of scientifically based evidence. "Policymakers are now eager and want to hear about even the modest effects that randomized trials may come out with," she said. "We've turned a corner, and there's more realism in the world of education. And a greater readiness." She ended her talk by declaring, "It's a whole new day."

Past and Current High School Reform Initiatives

From there, however, the focus shifted — not forward to dawning changes but back to the past. **Larry Cuban,** historian and Professor Emeritus of Education at Stanford University, sparked a debate that provided historical context for the rest of the colloquium when he presented his paper on the late twentieth century's repeatedly disappointing attempts to reform high schools. Arguing that rigorous, research-based evidence has never been used to improve the policies or practices behind high school reform, Cuban outlined the three main reasons he sees for the general failure of reform: the numerous conflicting purposes of high schools, the infrequent examination of historical evidence from early reform movements, and the implementation of policies without the use of available evidence to support or question those policies. (For further detail, see the full paper, included in this publication.)

Charles Payne, Sally Dalton Robinson Professor of African-American Studies, History, and Sociology at Duke University, took up several of Cuban's themes. He stated his belief that reform efforts have indeed led to changes but cautioned that, if they are to instigate real improvements, evaluations of reform initiatives must take individual schools' socioeconomic contexts into account. Finally, Cuban, Payne, and the session moderator, MDRC Senior Fellow **Jim Kemple,** took questions from the audience and hashed out the degree to which research has

made a difference historically, the ways in which education research might learn from other fields (such as the medical or public health models), and the factors that restrict schools' and districts' capacities to make use of research.

National, State, and Local Perspectives

The next panel examined reform from three unique but interrelated perspectives: the national, the state, and the local. Offering the national perspective was **Hans Meeder,** Deputy Assistant Secretary of the U.S. Department of Education's Office of Vocational and Adult Education, which has made high school reform one of its priorities and has scheduled a series of regional forums around the country in order to open up dialogue with states. "Experimental research will tell us things that other kinds of research cannot," Meeder said. But, he emphasized, a tension exists between the amount of research available at any given time and the pressure that policymakers face to make changes. The tension should prompt the education world to ask itself a couple of questions: Are there structures in place to implement research as it emerges? If not, how can we install them?

Consultant and state education analyst **Judy Bray** offered the state perspective, focusing the attention on one group in particular, the group that she said "reads the research": education policy analysts. She argued that policymakers are more likely to listen to analysts than to researchers. She also said that states are likely to become interested in the programs that other states are implementing. Researchers, she continued, should be aware of the many factors that influence policymakers' decisions, including state standards, exit tests, the requirements of the No Child Left Behind Act, the transfer of power between states and localities, school choice, and the fact that states tend to focus more on policies for early grades than for high school.

The next speaker, **Shirley Schwartz,** presented the local point of view, returning to the idea that analysts are an important element in education debates. Instead of saying, "Listen to us," Schwartz said, researchers should ask analysts, "What is it that you need to know?" And what they need to know, she argued, is the answers to questions such as: What are the findings on adolescent literacy? What are the research-based interventions that will work to improve high schools? And what is the best way of preparing new teachers for their jobs?

The Effects of High School Reform Initiatives

On the second day of the conference, the focus was on both the role of research and the various methods of studying high school reform. Presenters used slides to illustrate case studies and to delineate points about methodology. **David Stern,** Professor of Education at the University of California at Berkeley, presented a paper he co-wrote with **Jean Yonemura Wing,** which addressed the question of whether there is solid evidence of positive effects of high school reform.

(For details, see the full paper, included in this publication.) He mentioned that there is limited evidence from studies that meet high scientific standards, citing only three examples of programs that have undergone rigorous evaluation and produced evidence of positive effects: The Quantum Opportunities Program, Upward Bound, and Career Academies. Studies of other programs are often "informative" and even "ingenious," Stern said, but he used the example of studies on small schools to argue that those studies are problematic because they do not rely on random assignment: "They leave considerable room for doubt about the extent to which smaller school size causes better results for students," he said.[1] In response, **Richard Murnane,** Professor of Education and Society at Harvard University's Graduate School of Education, offered an economic perspective. Pointing out that small-school interventions are expensive, he suggested that collecting hard evidence about their effectiveness is all the more important.

The focus then turned to the issue of measuring the effectiveness of high school reform programs, with an emphasis on how to think about the magnitudes of the impacts that such reforms have on student behavior and performance — issues that are central to both designing evaluations and interpreting their results. **Howard Bloom,** Chief Social Scientist at MDRC, and **Mark Lipsey,** Director of the Center for Evaluation Research and Methodology at Vanderbilt University, presented a discussion of how to measure and interpret the magnitude of program impacts in general and of program impacts on educational outcomes in particular. They focused on the concept of effect size, a metric used widely in education research. Presenting a range of examples of program impacts reported in an effect-size metric, they compared these results with relevant external benchmarks (such as the average annual learning gain for students in schools or the average learning gap as measured by race/ethnicity) for judging how big a program effect must be in order for it to be considered important or policy-relevant.

But they also pointed out that effect size in different contexts can be interpreted in different ways. For example, Bloom cited Jacob Cohen and Mark Lipsey's "rules of thumb" for social science research, explaining that these rules are often used as a basis for judging the magnitudes of effects for high school programs: Roughly 0.2 standard deviation is considered a small effect size, 0.5 standard deviation a medium effect size, and 0.8 standard deviation a large effect size. (However, Bloom said, in a 1987 randomized clinical test of aspirin's ability to prevent heart attacks, an effect size of just 0.06 standard deviation led to a sweeping decision to end the study abruptly and prematurely. The decision was based on the feeling that aspirin's

---

[1]Studies that rely on random assignment use a lottery to divide students into groups that either do or do not have access to a given program. In such studies, comparisons between the two groups yield reliable estimates of the effect of the intervention in question. On the other hand, in studies that do not use random assignment, the same factors generating student selection into the program in question may generate inaccurate estimates of the program effects.

observed reduction in heart attacks was too large to continue giving a placebo to control group members. Bloom asked the audience to consider this example as a reminder that there are different ways to interpret effect-size findings when assessing an intervention or comparing the relative performance of alternate interventions.)

Studying Approaches to High School Reform

Fred Doolittle, Vice President of MDRC's Department of Education, Children, and Youth, introduced the next session, on methodology, with a list of some important rules of thumb for researchers to keep in mind as they design studies: Avoid a focus that is too narrow; create replicable studies with multiple sites; conceive of studies with long follow-up periods to allow for greater certainty; and identify intermediate benchmarks so that results can be meaningfully considered while a study is still in progress.

Presenters then described several case studies to demonstrate the strengths and the challenges of various research methodologies used to investigate school-level interventions, principles of school design, and district-level policy levers for change. Jim Kemple and MDRC Research Associate **Corinne Herlihy** offered results from the Career Academies and Talent Development evaluations;[2] **Barbara Means,** who directs SRI International's Center for Technology in Learning, and **David Rhodes,** Senior Research Analyst at the American Institutes for Research, discussed school design principles, based on their evaluation of the Gates Foundation National School District and Network Grants Program; and **G. Alfred Hess, Jr.**, Research Professor of Education and Social Policy at Northwestern University, presented a paper that explained how a district-level strategy for high school reform led to increases in student achievement in the City of Chicago. (For details, see the full paper, included in this publication.) At the end of the session, the Dean of Temple University's College of Education and former Senior Vice President of MDRC, **Kent McGuire,** commented that, together, the three presentations narrowed in on a complicated challenge: that high school reform interventions are "dynamic," requiring reconciliation of multiple interests and points of view while adhering to a need to maintain the substance and integrity of a solid research design. In the face of this, he said, the focus must be on obtaining valid and reliable measures for key outcomes and on drawing useful comparisons between outcomes for students served by the reforms and those for other, similar students.

Janet Quint, Senior Research Associate at MDRC, led the next session, introducing **Anthony Amato,** Superintendent of New Orleans Public Schools, and **Bernard Taylor, Jr.,**

---

[2]For the full text of MDRC's most recent reports on the Career Academies and Talent Development projects, see www.mdrc.org.

Superintendent of the Kansas City, Missouri, School District. She noted that administrators leading school reform efforts can bring distinctive and valuable perspectives to a research audience.

Taylor presented a pragmatic view of evaluations that rely on rigorous evidence, listing questions that he said "people want to know the answers to": Are kids doing better than they were the previous year? Are they graduating? Are more of them ready for college and a career? Where are we and how are we going to get elsewhere? Where do we want to go? What behaviors are associated with the types of programs that get you where you want to be?

Like Taylor, Amato focused on practical concerns. He described the often inadequate resources with which schools and districts must try to meet standards, arguing that research sometimes fails to take into account the limitations that specific schools might face (as an example, he described a hypothetical situation in which researchers attempt to install new technology in a high school without first assessing whether the school has the electrical capacity to support the technology). He summed up by listing three key research-related issues that superintendents face: the pressing need for information, the importance of testing reforms in real-world situations, and the need to focus at times on trends in student outcomes of special concern, rather than on the impact of specific reform efforts.

Reflections and Recommendations

The final session was a time for participants to reflect on the major issues discussed during the colloquium and ideas for what should happen next. David Ferrero of the Gates Foundation emphasized the need to find links among different approaches to knowledge, such as the research approach and the clinical one. **Steve Fleischman,** Principal Research Scientist at the American Institutes for Research, stressed the need for realism, asking which questions can be posed and answered and how research can build in a bigger focus on cost-benefit considerations. The Department of Education's Phoebe Cottingham, reiterating that there has been a change in education in the past twenty to thirty years, pointed to two new challenges: figuring out how to mesh the viewpoint of policymakers with that of practitioners and finding ways to embed the classroom perspective into the research.

Running through all the participants' summary comments was a major theme that reflects one of the driving forces behind the idea of the colloquium: In order for rigorous research to have a real impact on high school reform, the main groups involved in reform — researchers, policymakers, and practitioners — must communicate, working to integrate their very different perspectives into strategies that serve common goals. Other themes that emerged supported this one, and they also supported the idea that started off the colloquium during the introductory session: Education research has matured to a point where it potentially can have a very direct effect on policy and practice.

Key Themes

- **Context is critical.** As Charles Payne argued, researchers must take into account the physical, social, economic, and political situations of the schools and districts involved in their studies, and they must consider these things when they design studies, conduct evaluations, and develop conclusions. Without tests in multiple settings and without contextual knowledge, the results of evaluations may be misinterpreted. Furthermore, researchers, policymakers, and practitioners may have difficulty communicating with one another if they do not share the same contextual knowledge. This is because, with different bases of knowledge, there may be, as Payne put it, "a refusal to believe that a program supported by research can work here, even if it works elsewhere."

- **The evidence available so far from random assignment evaluations of high school reform is limited, but several new studies use the next-best alternative — a strong quasi-experimental design.** More random assignment studies on high schools are needed. However, several studies currently under way — with plans to produce results in the next year — use the next-best alternative: strong quasi-experimental designs (MDRC's Talent Development evaluation, combining interrupted time series analyses with comparison groups, is a good example). On a more abstract level, while it is important to acknowledge that research has the power to teach us empirically what will make a difference and what will not, we must also ask how much research is enough. David Stern emphasized that it's "unrealistic" to think there will ever be enough research to answer all questions definitively. Nevertheless, researchers and policymakers must look at what research is available and consider which programs have, in some ways, "worked."

- **Researchers need to challenge their own presumptions and premises.** Rather than making presumptions about practitioners' willingness to be involved in research or about their understanding of it, researchers should work to include the practitioner perspective from the beginning. This is especially important because practitioners frequently perceive a wide gulf between researchers' interests and findings and their own abilities to implement research findings in practical ways that make a difference in schools and classrooms. Without a link to operational realities and a partnership with groups that are able to make things happen, change will come slowly. This was a point that several participants emphasized when they offered their reflections at the end of the colloquium. Charles Payne brought it home when he handed paper to all participants and asked them to list, on one side, a project they had

undertaken that they had thought would be a "winner" but that had failed in some way, and to note on the other side what it was that they hadn't known at the time that had prevented the project from succeeding. This exercise, he said, helps explain "how smart people go wrong"; that is, how the presumptions people make can force them down the wrong roads.

- **Researchers must understand a variety of ways to describe the size of program impacts.** In their presentation on effect size, Howard Bloom and Mark Lipsey had one primary goal: to stimulate further thought about how to interpret what an effect-size measure means, not only with respect to interpreting the results of an impact study to assess the effectiveness of a given program but also in terms of planning for the sample size of a prospective evaluation study to ensure that it will provide enough statistical power to reflect a meaningful minimum detectable effect. In general, as Bloom and Lipsey and others made clear, researchers must take context into account when they analyze impacts. Also, given the challenge of using random assignment in education studies, researchers must find innovative ways of rigorously measuring key outcomes and analyzing data to make the most out of opportunities for impact studies.

- **Communication is of primary importance.** Researchers need to know how to make knowledge and research available to policymakers, school administrators, and teachers in a timely fashion. They also need to incorporate feedback mechanisms earlier so that practitioners and policymakers can learn about studies even if results are not yet available (perhaps making better use of the Web to this end). Because of the frustration that both researchers and practitioners feel when it comes to understanding each others' perspectives, it is important to strengthen relationships between the two groups so that the needs of each are more easily accommodated. On a related note, many conference participants noted the necessity of creating opportunities — such as this conference series — for connecting and networking with a range of people involved in high school reform.

- **Research evaluations should be longitudinal and hierarchical in design, and they should incorporate multiple questions and methodologies.** Fred Hess observed that studying a massive system like the Chicago school district requires a complex approach (indeed, the Chicago evaluation would have reached erroneous conclusions had it been a simple study of changes in outcomes over time). In order to be in a position to decide which programs are the best bets, as Fred Doolittle put it, it is crucial to design studies with longer follow-up periods. At the same time, studies should occur in multiple sites; they

should ask questions about implementation and effects or impacts; and, to produce the most reliable results possible, they should incorporate a number of methodologies. As a couple of participants noted: "It's the questions that drive the methodology"; and "every evaluation needs to involve multiple questions and multiple methodologies — both quantitative and qualitative."

- **There is an urgent need for additional information on the efficacy of small schools.** The Department of Education's Hans Meeder asked in his presentation, "Is small all?" The education community is currently focusing on small learning communities, small class sizes, and small schools as methods to improve high schools. While a small-school environment benefits many students in terms of their abilities to study and learn, small schools, in and of themselves, may not be sufficient to ensure increased student achievement, especially in new schools that are struggling to get off the ground. As Jim Kemple and others pointed out, large schools do offer some advantages, such as diversity in the student body, faculty, and curriculum. Small schools offer other clear advantages (such as more personalized and supportive learning environments), and more rigorous research is needed in order to answer questions about just how effective they are.

- **Both structured research and practitioner experience can generate knowledge, and too often one source of knowledge is privileged over the other.** Several researchers noted the difference between and varying strengths of "clinical knowledge" (the knowledge that comes from working in the field) and "research knowledge" (the knowledge that comes from evaluations and studies). For many at the meeting, the strongest evidence base draws on both types of knowledge. As noted earlier, David Ferrero addressed this question, stressing the need to see the links between these differing perceptions and to determine not whether clinical or research knowledge is more valuable but how both can be used to understand what a given school can accomplish in its specific context. Perhaps even more important, rather than viewing research knowledge and clinical knowledge as conflicting, the challenge is to bring the two together to complement and support one another.

- **Researchers should expect to face challenges when evaluations of high school reform programs involve random assignment.** It is important to keep in mind that random assignment is feasible only in certain situations, such as when the number of students demanding entrance to a small school or a charter school exceeds the available slots. While random assignment is a fair and equitable process, its use in the field of education is relatively new, and school administrators often have concerns or anxieties about using a lot-

tery-type process to select one group of students for an intervention and not another, or one group of schools and not another. It may take time for the random assignment design to become more widely respected in the education field and for school leaders to recognize that the reliability of the findings outweighs the inconveniences of participating in a study. On the whole, researchers need to shoulder the responsibility to make the rationale for random assignment — and the utility of it — clear to school personnel.

- **There is a need to acknowledge the limitations that affect research.** Tradition, politics, skepticism, lack of communication, lack of funding, and the difficulty of designing long-term, complex studies: Many factors inhibit the success of research studies. Many of them represent issues that can be resolved with clear, open communication and regular, communitywide conversations. But, on a pragmatic level, members of the education research community need to ask how they can work within the limitations: As Steve Fleischman put it, "Which questions can we ask and answer?" Along those lines, it is also necessary to consider cost-benefit issues and to carefully define the audience for each study.

Next Steps

While the colloquium built momentum for a dialogue among researchers, practitioners, and policymakers, there is much more work to be done. The next event in this conference series will be a two-day meeting in the spring or summer of 2005. The conference planners from this first gathering reconvened in April 2004 to discuss goals and directions, and they were joined by David Ferrero from the Gates Foundation, Jorge Ruiz-de-Velasco from the Hewlett Foundation, Angela Hernandez-Marshall (who represented Naomi Housman of the National High School Alliance), Barbara Means from SRI International, Steve Fleischman from AIR, Nancy Adelman from SRI, Monica Martinez from the Institute for Educational Leadership, Kent McGuire from Temple University, and several staff members from MDRC. Our challenge was to consider the perspectives of the three distinct audiences — researchers, policymakers, and practitioners — and to design a second event that will be relevant and meaningful for all three groups. The next conference will feature a more in-depth exploration of many of the issues laid out in this publication, with more time allotted for small-group discussions and dialogues about the different concerns of stakeholders. One particular focus will be the issue of how different questions lead to specific research methods; another will be the relative value of the findings. Also, the conference will include information from the comprehensive school reform studies, the results of which are scheduled for release early in 2005.

Why Has Frequent High School Reform
Since World War II
Produced Disappointing Results
Again, and Again, and Again?

Larry Cuban

Introduction

Because I will make a number of claims about the constancy of high school reform and its disappointing results in the past half-century, readers should know the basis for these statements beyond the footnotes I provide.

I have been an urban high school teacher for 14 years. I directed a school-based teacher education program in a Washington, DC, high school in the late 1960s, which became the model for the National Teacher Corps. I was the first director of a districtwide professional development program in the District of Columbia public schools in the early 1970s and subsequently served for seven years, until the early 1980s, as a superintendent in a small urban district with four high schools. As a trained historian for the past two decades, I have studied U.S. schools in the nineteenth and twentieth centuries and the repeated attempts to reform the governance, organization, curriculum, and instruction in these institutions.

Drawing from this research-based and experience-produced knowledge, I make three statements encompassing past and current high school reforms and the research surrounding these efforts to answer the puzzling question posed in the title of this paper. Each statement I will elaborate enough to give flavor and substance.

Past and Current High School Reforms

Statement 1: Promoters of high school reform have largely ignored the historical evidence of that institution's conflicting purposes.

I begin with the very inception of tax-supported public schools and the "common school" ideal articulated by mid-nineteenth-century reformers like Horace Mann. They sought social and political harmony to calm the growing social friction in urban and industrializing New England and mid-Atlantic states between the landed wealthy, an emerging merchant class, skilled artisans, and unskilled laborers — mostly immigrant and displaced rural poor. Creating schools for all of a community's children would bring different classes together and solve a growing societal problem. Historian Lawrence Cremin crisply summarized the ideal:

> The common school was to be for rich and poor alike, not only free but the equivalent in quality of any comparable institution. In it would mix the children of all creeds, classes, and backgrounds, the warm association of child-

hood kindling a spirit of mutual amity and respect, which the strains and cleavages of adult life could never destroy.[1]

The common-school ideal guided the establishing of tax-supported grammar schools, yet a small but growing middle class also wanted its sons and daughters to enter college. In Massachusetts, for example, mostly middle class voters established tax-supported high schools as early as the 1820s. Open to all families in a community — from the children of wealthy landowners and successful merchants to those of farmers and laborers — the small number of high school students who passed the exacting entrance examinations took rigorous academic subjects, including foreign languages. Most never completed the four-year curriculum, leaving high school to enter business as clerks and bookkeepers or to teach in elementary schools.[2]

High schools slowly spread after the Civil War to accommodate more upper- and middle-class children (although small percentages of working-class children did attend), but the "people's colleges" remained small and scattered in small towns and clustered in cities. Even by the 1870s, less than 4 percent of all students were enrolled in high schools. The average urban high school in these years had only 85 students and 3 teachers. Less than one-third of those who passed the entrance exams and enrolled ever gained a diploma. The high school, as historian David Tyack summed up its purpose, was to be "an academically elite school fit for the first families, but it was not to be restricted by social class."[3]

These selective institutions prided themselves on being both democratic and meritocratic. One of the founders of Philadelphia's Central High School (1839) said years after the establishment of the school:

> It is the School of the Republic . . . opening its portals alike to the son of a President or a ploughman, a Governor or his groom, a millionaire or a hewer of wood — treating with equal justice — rearing with equal fidelity, and crowning with all its honors alike the one and the other, and demanding no passport to its blessings, or to its laurels, save that which the people demand, and forever will demand from all its sons — INDIVIDUAL, PERSONAL MERIT. Such, fellow citizens, is your High School.[4]

Intended to be democratic and merit-based, the original purposes of the high school were to create an aristocracy of academic excellence from children of all social classes. Yet

---

[1]Cremin, 1957, p. 8. During Cremin's time, except for rare instances, blacks and other minorities were excluded from the common school. Also see Wraga, 1994.

[2]Reese, 1995; Katz, 1968.

[3]Tyack and Hansot, 1990, pp. 117, 123.

[4]Labaree, 1992, p. 26.

these high schools were also to prepare youth for college and the workplace if they left before graduating. Being democratic — provide access to all — meritocratic — only the best graduate — and practical — preparing all students for jobs, business, and the professions — have been deeply embedded and competing purposes for the high school ever since.

Achieving all these purposes in an institution having a single academic curriculum as more and more students attended proved to be especially difficult in the twentieth century, until progressives invented the comprehensive high school.

Tensions over reaching these competing purposes escalated greatly in the decades following World War I as more students entered high school and it became clear that most of them would not go on to college. In 1900, one in ten youth ages 14 to 17 were enrolled in high school; by 1940, seven of ten were.[5]

To ease these strains, progressive educators, thinking that they knew the vocational futures of high school students, created multiple curricula for those preparing for college, those who entered commercial jobs, those who sought industrial work, and others uncertain of their work futures. Using large group tests developed during World War I, progressives determined which students would be placed in which curriculum. In addition, sports programs, clubs, newspapers, proms, graduation, and other activities created a virtual community of adolescents. All of the formal and informal accoutrements of comprehensive high schools helped reconcile the growing tensions — at least for a while — between keeping a high school democratic, meritocratic, and practical.

Yet the comprehensive high school, like most social institutions, was hardly insulated from political, economic, and social problems arising from larger changes in the culture. The progressives who created the comprehensive high school also wanted all youth to learn about the temperate use of alcohol, the evils of smoking tobacco, and, when automobiles became universal, safe-driving skills to reduce highway accidents and deaths. Subsequent generations of reformers drafted both elementary and secondary schools into helping the country fight foreign enemies in two world wars and a cold war. Reformers in the 1960s sought social justice by eliminating segregated schools — often beginning by desegregating high schools — and reducing the huge learning gaps between poor minority and white middle-class children.

The dominant reforms adopted and implemented for the past two decades offer another example of national and state policymakers turning to schools to remedy social and economic ills. Contemporary policymakers have grasped the comprehensive high school — itself a product of an earlier reform movement — with its diverse curricula, menu of electives, and exten-

---

[5]Tyack and Cuban, 1995, p. 47.

sive extracurricular program, and converted it into a leaner, uniformly academic, and more ac-countable institution resembling the late-nineteenth-century high school in its college-preparatory concentration. Why? The overriding rationale has been to align the "new" high school to an ever-changing global economy where employers would be able to hire youth from either rich or poor backgrounds with the skills and knowledge to go directly to college or to en-ter the workplace after graduation. No more mismatch between what employers need in an in-formation-based, highly competitive economy and what graduates offer with their skills fitted to an industrially based economy. Thus, to regain global economic supremacy that had been lost in the 1970s and 1980s and to reassure U.S. taxpayers, employers, and the larger society, the con-temporary high school has had to be closely tied to a changing economy.

If there has been one constant refrain in U.S. reform, then, it has been the reliance on the common-school ideal, and later on the comprehensive high school, as a cure for national ills. In 1898, the Boston superintendent of schools told the National Education Association that pub-lic schools were "the salvation of the American Republic." Not recorded then were the "amens" greeting the superintendent's statement.[6]

In helping to solve national problems, the spread of high schools in the twentieth cen-tury and the growing importance of the high school diploma as a passport to a well-paying job particularly strengthened the personal expectation most parents held about completing 12 years of schooling: their children would receive a ticket that placed them on a socioeconomic escala-tor, lifting graduates from immigrant, low-income, and working-class families into the middle and upper classes. Seldom made explicit as a public purpose for tax-supported high schools, the self-interested (and private) goal parents held for sending their sons and daughters to secondary schools was to increase their children's chances of securing financial and social success.

Beyond reformers' seeing public schools as tools for solving national problems and par-ents' seeing schools as vehicles for their children's success, other reform-minded academics, poli-cymakers, and practitioners challenged the dominant view of comprehensive high schools, with their differentiated curricula and tracking of students into particular classes, as the best way to educate youth for the twentieth century and beyond. Some progressive reformers in the 1920s and 1930s (and since) envisioned the high school as a humane, intellectually challenging, community-oriented, and democratic place where the young could think critically and independently, appreci-ate productive labor, and build a just society. They wanted students to enhance their personal well-being while in high school and become proficient in whatever unique strengths they possessed.

---

[6]Lazerson, 1971, p. 202. Many other historians and social scientists have noted the frequency of reform-ers' turning to schools to solve national ills: See Katz, 1987, pp. 111-135; Cremin, 1990, pp. 85-127; Kaestle, 1990; Tyack and Cuban, 1995, pp. 40-59. Also, for a discussion of how interest groups in a democracy foist goals upon public schools, see Chubb and Moe, 1990, pp. 53-55. For more general information about high schools, see Krug, 1964; and Wraga, 1999, pp. 523-544.

High schools were to be communities where students would live for four years and experience intellectually ambitious teaching, close working relationships with adults in and out of school, and collective decision-making about issues inside and outside the school.[7]

Juggling the varied democratic, meritocratic, and practical purposes embedded in the history of high schools along with the private aspirations of parents for their children and the insistent reform impulse to use high schools as solutions to national problems — all within the framework of the comprehensive high school — has produced severe strains among reformers and confusion among parents, especially since World War II. Writers have often reduced these conflicting and often implicit purposes to labels of "conservative" and "progressive."[8] While what happens in schools is far more complex and nuanced than these labels allow, the pigeon-holing does mirror the shorthand language that policymakers, media, and elected officials use in public discussions of school reform. Moreover, today, reformers still promote one or more of these purposes in efforts to reshape the comprehensive high school to make it conform to their visions of how the high school and the larger society ought to be. So I use these terms, reductionist as they are.[9]

Consider the late-nineteenth-century and early-twentieth-century college-preparatory high school as an incarnation of the conservative purpose. Past and present, public and private, such schools as Boston Latin, Bronx School of Science, Lowell High School (in San Francisco) — and newer versions such as Downtown Prep, a charter school in San Jose — testify to the continuing strength of highly competitive and selective schools in preparing the young from all backgrounds for college and for eventual high-paying professional, business, and managerial posts in the workplace and in the civic arena.

Also consider other kinds of high schools, not as prevalent as traditional college-preparatory types, but seeking the progressive purposes of a schooling that is intellectually stimulating, has cross-disciplinary curricula, and promotes personal well-being by creating communities where sustained relationships between teachers and students are salient in working to improve both the school community and the larger society. For example, there are currently community-based, democratically governed, and social-action-oriented high schools, such as the Robert F. Kennedy Community High School in Queens, New York City; Hanover High

---

[7]Cuban, 2003; Rury, 2002.

[8]I use these commonplace terms in their educational — not political, partisan — sense. I attach no negative or positive values to either.

[9]For an excellent analysis of multiple and conflicting goals for U.S. public schools, see Labaree, 1997.

School in Hanover, New Hampshire; and Metropolitan Regional, Technical, and Academic High School in Providence, Rhode Island.[10]

Although I have presented these purposes and the examples of schools as mutually exclusive, there is clearly overlap in practice between and among them, including many hybrids of both types of schools. A college-preparatory high school in which 99 percent of the students attend four-year universities, for example, can be a humane community where cooperation and community service are prized. The overlap in particular features or hybrids, however, should not obscure the deep ideological differences in overall social purposes for these schools and the warlike language that has historically divided partisans of each kind of high school. It was precisely this frequent rhetorical and policy conflict over the competing purposes of the high school (that is, democratic, meritocratic, and practical) that drove an earlier generation of reformers to invent the comprehensive high school in the early twentieth century to reconcile divergent, even conflicting purposes.

Although, by the 1950s, the comprehensive high school had become the dominant form of secondary school organization in the United States, demographic, social, and economic changes have led both conservative and progressive critics of the high school to find much fault with these institutions. This brings me to my second statement.

Statement 2: Since World War II, high school reformers seldom examine or use the historical evidence of earlier movements to change comprehensive high schools.

Attacked constantly since the 1950s by conservative reformers for being anti-intellectual and watering down the academic curriculum, and attacked by progressives for being far too big and racially isolated and for reproducing the inequities in the larger society, the comprehensive high school has been a surprisingly resilient institution. Even today, with its 1,500-plus students, a full range of after-school activities, and a curriculum that is increasingly college preparatory (with reduced commercial and vocational offerings), the comprehensive high school catering to the vast majority of students continues to generate and receive scorching censure. A former assistant secretary of education in the Clinton administration spared few words in 2001 when he said: "At a time when high schools must be pathways to college for all students, they are pathways to nowhere for many."[11]

---

[10]For Hanover High School, see Mosher, Kenny, and Garrod, 1994, pp. 5-21; for Robert F. Kennedy Community High School, see www.insideschools.org/view/ed_newqueensHS; for the Metropolitan Regional High School, see www.metcenter.org.

[11]Cohen, 2001, p. 3; Rury, 2002.

One constant criticism has been the enormous variation among comprehensive high schools. For example, within the city of San Diego, a LaJolla High School and San Diego High School differ enormously in daily culture, quality of teaching, and student outcomes. Across districts, the same divergence occurs, say, between New Trier High School in Evanston, Illinois, and nearby Wendell Phillips High School in Chicago. One cannot ignore the commonsense observation that race and class play a large part in shaping routines, cultures, and student outcomes in comprehensive high schools. Unyielding criticism of comprehensive high schools from conservative and progressive reformers preceded wave after wave of innovations, especially after World War II.

The economically and politically driven school reforms themselves, most often fueled by external events rather than by research studies, convinced policymakers to make changes in high school governance, organization, curriculum, and instruction. One has only to mention Sputnik in the 1950s Cold War era, the Civil Rights Movement in the 1960s, and the market penetration of Japanese automobiles and electronics in the late 1970s to recall the high-pitched criticism of failing high schools and the swift, if feverish, adoption of such reforms as Advanced Placement courses, desegregated classrooms, site-based management, restructured schools, and higher graduation standards. None of these reforms was adopted because of accumulated evidence or persuasive research studies. In each instance, educational policymakers and administrators responded to the barrage of criticism of comprehensive high schools and successive waves of reforms by totally embracing some changes, selectively adapting others, and ignoring the rest.

Reformers in the 1950s sought to put a stainless steel spine in the curriculum. Fearing Soviet strength in space exploration, they wanted far more academic preparation and an increased number of students going on to college to become scientists, engineers, and mathematicians. School reformers steered comprehensive high schools toward traditional college preparation to be more aligned with national foreign policy objectives. Advanced Placement courses were installed and more math and science were added to the curriculum.[12]

Then the Civil Rights movement in the 1960s spilled over high schools as reformers sought ways of making the institutions more humane, egalitarian, and responsive to social injustices. Concerns over poor academic performance leading to dropouts and dead-end jobs mobilized reformers to desegregate high schools, introduce new programs that helped students move into college-preparatory courses, and created schools-within-schools and restructured alternatives that put into comprehensive high schools more progressive practices.[13]

---

[12]Ravitch, 1983; Powell, Farrar, and Cohen, 1985.
[13]Patterson, 2001; Kluger, 1977.

In the 1970s business and civic leaders pressed school policymakers to do something about the mediocre performance of U.S. high school students on international achievement tests and the inferior quality of entry-level workers in a rapidly changing workplace, especially workers drawn from urban schools. Beginning in the early 1980s and continuing to the present, corporate-inspired reforms moved comprehensive high schools to raise graduation standards, require students to take more academic courses, develop content and performance standards, and hold school staff and students responsible for improving academic achievement.

In these years, progressive reformers have not sat on their hands. They, too, critiquing large comprehensive high schools, pushed for alternative assessments, restructured programs, and smaller urban high schools. And, as before, educators adopted and then adapted changes in comprehensive high schools.[14]

If externally driven reform aimed at solving national problems — far more than research studies — have largely determined the direction of changes in comprehensive high school governance, curriculum, and organization, it is only since the 1970s that researchers have begun to concentrate on linking the multiple and conflicting purposes of the high schools to student outcomes rather than to the amount of resources spent on schools.

From the late 1980s through the 1990s, presidents, governors, mayors, and legislatures — often spurred by business leaders — began crafting reforms that sought to turn comprehensive high schools, including urban ones, into college-preparatory institutions. Vocational education courses preparing students for current jobs in the workplace shrank. In fact, the academic course of study — four years of English; three years of history, math, and science — has become a virtual vocational course of study aimed at college preparation.[15]

While the focus seemed to be on all high schools, most suburban and urban comprehensive ones that were already scoring high on standardized achievement tests and sending 80 percent or more of their graduates to four-year colleges (for example, Lowell High School in San Francisco) seldom had to cope with fiery groups of parents eager to reform the school.

Far more pressure was applied to urban comprehensive high schools that generated most of the scary articles on violence in schools, dropouts, gang warfare, drugs, and crime. Deep concerns on the part of the minority community for the future of their youth, business leaders projecting labor market needs in the next century, and civic officials wanting to restore social stability and commercial vitality to economically depressed areas of their cities joined forces to make changes in urban comprehensive high schools. The foundation-funded small-schools movement

---

[14]Toch, 1991.
[15]National Center for Education Statistics, 2001.

and preparing all urban youth for college — an echo of the late-nineteenth-century high school — have been popular reforms. Still, the urban comprehensive high school persists.

Researchers studied these developments after they occurred and reported on them but were bit players in the reform dramas that unfolded in the post-World War II decades. As these externally driven reforms swept across the high school landscape and reformers leaped at one innovation after another, the larger relationship between the conflicting purposes of high schools and their effects on students went largely unexplored, save for a few uncommon efforts. This brings me to my third statement.

Statement 3: For the most part, high school reformers, in adopting and implementing policies, seldom used available evidence (including historical investigations) to support or question the reforms or their outcomes.

Assailed as it has been from both conservatives and progressives, the comprehensive high school is not the unalloyed failure that critics say it is. Suburban comprehensive high schools, drawing from middle-class and upper-middle-class families, have displayed high test scores and high rates of college attendance, low dropout rates, and few disciplinary actions with students. Moreover, these schools have easily adapted to the reforms engineered in the 1980s and since.

Nowhere near perfect, these schools still draw criticism for their frenzied competition among students and high rates of drug and alcohol use, and for ignoring low achievers in their midst. Nonetheless, these schools rank high in books and magazines listing the 100 best high schools in the nation.[16]

While the small-schools movement has found some suburban schools hospitable, most of the action for small high schools has occurred in cities. Certainly, the majority of (but by no means all) urban comprehensive high schools have high rates of dropouts, low academic achievement, low attendance, and small percentages of graduates enrolling in colleges. These failing urban schools have become sites for the small-schools movement.[17]

So, currently, there is a split verdict on the worth of the comprehensive high school. Anxious middle-class families spend much money to buy homes in suburban districts where comprehensive high schools score well on state and national tests and high percentages of students enter four-year colleges. If school boards and superintendents want parent rebellions on their hands, they could try converting Pittsburgh, Pennsylvania's Taylor Allderdice High

---

[16]Mathews, 1998; Pope, 2002.
[17]Cohen, 2001.

School, or Montgomery County, Maryland's Walt Whitman High School, or California's Beverly Hills High School into buildings housing clusters of small schools (including chartered ones) with different programs; or they could try supporting vouchers for religious schools. Yet in most big cities, in high schools where dropout rates cut senior classes to a fraction of what they were in the ninth grade and where daily attendance is low, small academic schools, charters — and in some instances even vouchers — are welcomed, if polls are to be believed.

Where too many (but not all) researchers have been timid in the past century has been in laying out for policymakers (particularly when they are the funders of and the audience for the research) the chain of logic and evidence supporting critical policy assumptions. Were a reader to raise an eyebrow over my questioning researchers and their relation to policymakers in doing studies that are policy-relevant, consider the serious problems of assessing the worth of current reforms that face business and civic leaders, parents, media editors, and practitioners who continue to believe that standards-based curricula, tests, and strong accountability measures will yield improved student test scores and a reduction in the achievement gap between whites and underperforming minorities.

First, one would have to determine whether widely praised national, state, and district strategies and practices actually produced higher test scores, and, second, one would have to verify that the scores are accurate measures of school success and that they predict future achievements.

The causal chain of policy logic runs as follows: Were state-mandated curriculum standards implemented as intended and, if they were, did the standards (and any infrastructure put into place) influence teaching practices? Did teaching practices, in turn, shape what students learned as measured by the state tests? The first causal linkage requires evidence that state policies (and the mechanisms supporting schools and teachers) were fully put into practice, thereby affecting classroom practices. The second shows that changed instructional practices resulted in desired outcomes. Even here, test score gains require scrutiny of the contribution that classroom experiences made to student achievement over a specific period of time, once controls for prior test performance and students' socioeconomic status are held constant.[18]

The second issue is whether test scores do indeed measure current and future success. Some researchers have raised questions about whether standardized achievement test scores are proxy measures of school productivity.[19] The many purposes of comprehensive high schools (democratic, meritocratic, practical, and solution to national problems) and the varied cultures within schools make single, quantifiable measures of success dubious.

---

[18]For this point, I found Porter and Smithson, 2001, most helpful in my thinking.

[19]For test scores, see Levin, 2001; Hanushek, 1994; and Balfanz, 1991; for links between high school performance and workforce productivity, see Smith, 1999.

The difficulties of measuring success are not confined to schools. All nonprofits (and profit-making organizations as well, but that is for another paper) share issues of measuring outcomes consistent with their primary goals. When a mental hospital's effectiveness, for example, is measured by how many beds are occupied, as Peter Drucker put it, "this yardstick leads to mental patients being kept in the hospital — which, therapeutically, is about the worst thing that can be done to them." Drucker raises the same issue in determining whether universities are successful. He asks which of the following are measures of "doing a good job": the salaries of students 20 years after graduation? the reputation of the faculty? the number of Ph.Ds? scientific prizes alumni have earned? "Each yardstick," Drucker points out, "bespeaks a value judgment regarding the purpose of the university — and a very narrow one at that. Even if these were the right objectives, such yardsticks measure performance just as dubiously as the count of bed utilization measures performance in mental hospitals."[20]

For reformers dead-set on using test scores, these two issues (that is, Were standards fully implemented? Did implemented standards change teaching and learning and yield desired student outcomes?) are essential in demonstrating the effects of policy changes and the impact of those changes upon instruction and, ultimately, upon student achievement in high schools. Thus far, researchers, policymakers, and business and civic elites have failed to take up these issues in public and the causal policy logic remains hidden or relegated to academic journals and gadflies.[21]

Finally, most researchers interested in reforming high schools have yet to demonstrate through rigorous studies that different types of high schools — progressive, conservative, or hybrid — have yielded the student outcomes they promoted.

Considering the Different Types of High Schools

The current model of a good college-preparatory school, large or small, is one that has high test scores, high rates of college attendance, and low dropout rates. It is far and away the dominant model in its monopoly on goodness and is unquestioned by the public, most policymakers, the media, and quite a few researchers. It is taken for granted because it seemingly ful-

---

[20]Kanter and Summers, 1987, p. 156. Also, Murnane (1987) analyzes similarities and differences in economic and educational indicators. He examines generic commonalties between the two in measuring unemployment and educational performance but recognizes two key differences. First, local school boards are responsible for raising children's achievement, while local employers and government officials are not responsible for local unemployment. Second, the decentralized governance of schools heavily influences the choice and quality of indicators, an influence missing from indicators of economic performance (pp. 101-116). For more comparisons of productivity between private and public sectors, see Triplett and Bosworth, 2000.

[21]A small number of researchers are exceptions to this statement. See, for example, Elmore and McLaughlin, 1988; Cohen, 1989.

fills the democratic, meritocratic, and practical purposes that conservative reformers say the high school should serve. Yet few researchers, if any, can state with confidence that the college-preparatory program and its teachers add a clear increment of value over what families contribute to their children's lives and what individual student motivation provides. Moreover, researchers looking at these large and small college-preparatory high schools have compiled studies that trumpet the effectiveness of these programs. The same issue of selection bias taints both kinds of studies. Neither set of researchers, however, seems to have swayed the other, although both say that *their* studies are rigorous, scientifically done, and so on.[22]

For a number of reasons, matched experiments that try to reduce selection bias — conducted by organizations and researchers who seemingly have no obvious or even covert agenda to prove one purpose or another is superior — have seldom tested, over an extended period of time, whether conservative and progressive purposes embedded in high schools yield the desired effects on students. Do students who attended these high schools graduate from college and perform well in jobs? Are graduates of these high schools civically engaged? In short, have these schools achieved their divergent purposes? I know of only a few studies (there may be others) that have even tried to investigate these outcomes:

- **The Eight-Year Study (1934-1941).** Privately funded by foundations to determine whether 30 high schools could design and implement progressive programs and whether, in college, graduates of those high schools would do as well as or better than graduates of typical high schools. Nearly 300 colleges cooperated by suspending their normal admission requirements; 1,475 matched pairs of students were studied. Graduates from progressive programs did as well as those from traditional ones.[23]

- **Follow-Through Evaluation (1970s).** Federally funded planned variation study that sampled and examined different models of preschool education. Graduates of the preschool programs under study were matched with nonpreschoolers and followed into their adolescence to determine whether differences accrued to preschool-educated children in academic and social outcomes. While design and methodological flaws were evident in the study and in the subsequent follow-up inquiries, it remains an ambitious and startling study in asking fundamental questions about different types of educational programs compared with present-day menus of studies offered by research-

---

[22]Mathews, 1998; Bensman and Meier, 2000.
[23]Aikin, 1942.

ers. The results, flawed as they were, again revealed that no one type of schooling is best for all young children.[24]

Conclusion

Readers may have already guessed that I am not optimistic that, were ambitious longitudinal studies like the two described in the previous section to be done today — and done well — and were those studies to show that there is no best way to school a 17-year old, policymakers would be stunned and would rush to embrace the results. Political, economic, and social changes in our society, leading to mobilizing coalitions to reform schools — not effect sizes of studies or the lure of experimental designs — prompt reformers to pursue high school reform along the paths I have suggested earlier. The belief that there is only one kind of good high school is deeply planted in their psyches, and that kind of school happens to be close to the traditional college-prep model. Few research studies have challenged this erroneous belief in a One Best High School.[25]

Rigorous impartial studies can establish a potential basis for policy action about more than one kind of "good" school and about alternative assessments that are far more sensitive than current standardized tests to what happens in high schools and classrooms. What now exists among researchers seldom goes beyond dueling studies put out by partisans of the different and conflicting purposes (democratic, meritocratic, and practical) deeply embedded in the history of U.S. high schools.

Here, then, is my answer to the question I stated at the beginning of this paper: Why has frequent high school reform since World War II produced disappointing results again, and again, and again? Historically, political, social, and economic changes in the nation have produced problems that have shaped school reformers' agendas. Research studies have played a minor-to-trivial role in either shaping those policy agendas or offering solutions. Moreover, most researchers have seldom examined the policy logic for its internal coherence and causal linkages to implementation in schools and classrooms. Reformers — both conservative and progressive — along with inspired allies outside of education have sought to use high schools as solutions for national problems. Finally, most reformers and researchers have overlooked, neglected, or forgotten (or perhaps all three) the conflicting purposes deeply embedded in the origins and growth of U.S. high schools.

For all these reasons, my answer to the question guiding this session — "How Far Have We Come and Where Are We Headed?" — is straightforward: At no point in the history of

---

[24]Kennedy, 1978, pp. 3-11.
[25]Ravitch, 1984, 2000.

high school reform, past and present, was rigorous research-produced evidence used to improve either policy or practice, nor do I expect that situation to change in the next decade.

References

Aikin, Wilford. 1942. *Adventures in American Education: The Story of the Eight-Year Study.* Vol. 1. New York: Harper and Brothers.

Balfanz, Robert. 1991. "Local Knowledge, Academic Skills, and Individual Productivity: An Alternative View." *Educational Policy* 5 (4): 343-370.

Bensman, David, and Deborah Meier. 2000. *Central Park East and Its Graduates: Learning by Heart.* New York: Teachers College Press.

Chubb, John, and Terry Moe. 1990. *Politics, Markets, and America's Schools.* Washington, DC: Brookings Institution Press.

Cohen, D.K. 1989. "Teaching practice: Plus Que Ça Change...." In P. W. Jackson (Ed.), *Contributing to Educational Change: Perspectives on Research and Practice.* Berkeley, CA: McCutchan Publishing Corporation.

Cohen, Michael. 2001. *Transforming the American High School.* Aspen, CO: Aspen Institute.

Cremin, Lawrence (Ed.). 1957. *The Republic and the School: Horace Mann and the Education of Free Men.* New York: Teachers College Press.

Cremin, Lawrence. 1990. *Popular Education and Its Discontents.* New York: Harper and Row.

Cuban, Larry. 1998. "The Media and Polls in Education." In Gene Maeroff (Ed.), *Imaging Education: The Media and Schools in America.* New York: Teachers College Press.

Cuban, Larry. 2003. *Why Is It So Hard to Get Good Schools?* New York: Teachers College Press.

Elmore, Richard F., and Milbrey McLaughlin. 1988. *Steady Work: Policy, Practice, and the Reform of American Education.* Santa Monica, California: Rand Corporation.

Hanushek, Eric A. 1994. *Making Schools Work: Improving Performance and Controlling Costs.* New York: Brookings Institution.

Kaestle, Carl. 1990. "The Public Schools and the Public Mood." *American Heritage* 41 (February): 61-81.

Kanter, Rosabeth, and David Summers. 1987. "Doing Well While Doing Good: Dilemmas of Performance Measurement in Nonprofit Organizations and the Need for a Multiple-Constituency Approach." In Walter Powell (Ed.), *The Nonprofit Sector: A Research Handbook.* New Haven, CT: Yale University Press.

Katz, Michael. 1968. *The Irony of Early School Reform.* Cambridge, MA: Harvard University Press.

Katz, Michael. 1987. *Reconstructing American Education.* Cambridge, MA: Harvard University Press.

Kennedy, Mary. 1978. "Findings from the Follow Through Planned Variation Study." *Educational Researcher* 7 (6): 3-11.

Kluger, Richard. 1977. *Simple Justice: The History of Brown v. Board of Education and Black America's Struggle for Equality.* New York: Vintage Books.

Krug, Edward, 1964. *The Shaping of the American High School.* Vol. 1. New York: Harper and Row.

Labaree, David F. 1992. *The Making of an American High School: The Credentials Market and the Central High School of Philadelphia, 1838-1939.* New Haven, CT: Yale University Press.

Labaree, David F. 1997. "Private Goods, Public Goods: The American Struggle Over Educational Goals." *American Educational Research Journal* 34 (1): 39-81.

Lazerson, Marvin. 1971. *Origins of the Urban School: Public Education in Massachusetts, 1870-1915.* Cambridge, MA: Harvard University Press.

Levin, Henry. 2001. "High-Stakes Testing and Economic Productivity." In Gary Orfield and Mindy L. Kornhaber (Eds.), *Raising Standards or Raising Barriers? Inequality and High-Stakes Testing in Public Education.* New York: New Century Foundation Press.

Mathews, Jay. 1998. *Class Struggle: What's Wrong (and Right) with America's Best Public High Schools.* New York: Times Books.

Mosher, Ralph, Robert Kenny, Jr., and Andrew Garrod. 1994. *Preparing for Citizenship: Teaching Youth to Live Democratically.* Westport, CT: Praeger.

Murnane, Richard. 1987. "Improving Education Indicators and Economic Indicators: The Same Problems." *Educational Evaluation and Policy Analysis* 9 (2): 101-116.

National Center for Education Statistics. 2001. "Stats in Brief: Changes in High School Vocational Coursetaking in a Larger Perspective." Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Patterson, James T. 2001. *Brown v. Board of Education: A Civil Rights Milestone and Its Troubled Legacy.* New York: Oxford University Press.

Pope, Denise. 2002. *Doing School.* New Haven, CT: Yale University Press.

Porter, A. C., and J. L. Smithson. 2001. "Are Content Standards Being Implemented in the Classroom? A Methodology and Some Tentative Answers." In Susan Fuhrman (Ed.), *From the Capitol to the Classroom: Standards-Based Reform in the States.* Chicago: University of Chicago Press.

Powell, Arthur G., Eleanor Farrar, and David K. Cohen. 1985. *The Shopping Mall High School: Winners and Losers in the Educational Marketplace.* Boston: Houghton Mifflin.

Ravitch, Diane. 1984. "A Good School." *American Scholar* 53 (4): 481.

Ravitch, Diane. 2000. *Left Back: A Century of Battles over School Reform.* New York: Simon and Schuster.

Ravitch, Diane. 1985. *The Troubled Crusade: American Education, 1945-1980.* New York: Basic Books.

Reese, William. 1995. *The Origins of the American High School.* New Haven, CT: Yale University Press.

Rury, John. 2002. "Democracy's High School? Social Change and American Secondary Education in the Post-Conant Era." *American Educational Research Journal* 39 (2): 307-336.

Smith, J. P. 1999. Tracking the Mathematics of Automobile Production: Are Schools Failing to Prepare Students for Work? *American Educational Research Journal* 36 (4): 835-878.

Toch, Thomas. 1991. *In The Name of Excellence.* New York: Oxford University Press.

Triplett, Jack, and Barry Bosworth, 2000. "Productivity in the Services Sector." Paper delivered at the American Economic Association conference, Boston, January 9.

Tyack, David, and Larry Cuban. 1995. *Tinkering Toward Utopia.* Cambridge, MA: Harvard University Press.

Tyack, David, and Elisabeth Hansot. 1990. *Learning Together: A History of Coeducation in American Public Schools.* New Haven, CT: Yale University Press.

Wraga, William. 1994. *Democracy's High School: The Comprehensive High School and Educational Reform in the United States.* Lanham, MD: University Press of America.

Wraga, William. 1999. "The Progressive Vision of General Education and the Common School Ideal: Implications for Curriculum Policy, Practice, and Theory." *Journal of Curriculum Studies* 31 (5): 523-544.

www.insideschools.org/view/ed_newqueensHS.

www.metcenter.org.

Is There Solid Evidence of Positive Effects for High School Students?

David Stern
Jean Yonemura Wing

Introduction

In this paper we illustrate the use of a strict standard for evaluating evidence on programs and strategies designed to improve outcomes for high school students. We explain what we mean by solid evidence, and present examples from multisite evaluations of three programs. After that we examine some of the evidence on high school size and explain why clear inferences about cause and effect remain elusive. We also look at examples of studies that use data for the whole student population in large districts, as an approach to reduce possible selection bias. We conclude with a predictable recommendation for more rigorous evaluation, and a programmatic suggestion.

Although we concur that random assignment studies provide the best support for inferences about cause and effect, we reject the idea that random assignment should be the method of all or most educational research. Before a program, strategy, or intervention can be tested by random assignment, it has to be formulated from exploratory research, and tried out in nonexperimental settings. History, ethnography, case study, design study, and other kinds of research all contribute to the understanding of educational phenomena and the development of new ideas. Even at the final stage of testing the effectiveness of a particular intervention, qualitative information about the experience of participants is useful in suggesting why effects do or do not occur and how the program or strategy might be further improved. Also, as mentioned below, some hypotheses do not lend themselves to testing by random assignment.

Cautionary Tales

"Over the past 30 years, more than two dozen comparison-group studies have found hormone replacement therapy for postmenopausal women to be effective in reducing the women's risk of coronary heart disease, by about 35-50 percent. But when hormone replacement therapy was finally evaluated in two large-scale randomized controlled trials — medicine's 'gold standard' — it was actually found to do the opposite: it *increased* the risk of heart disease, as well as stroke and breast cancer" (emphasis in original).[1,2]

Education, more than modern medicine, is notoriously susceptible to fads. Remember the school-to-work movement? At one high-level meeting in the mid-1990s, the front of the conference folder had "The School-to-Work Movement" printed on stick-on labels that were not quite firmly attached. The labels covered what had first been printed on the folder by mistake: "The School-to-Work Moment." In retrospect, this was probably a more accurate title.

---

[1]U.S. Department of Education, 2003, section I.C.2.

[2]It is important to note that these random assignment trials tested effects of a combined dose of estrogen and progestin, but in practice estrogen is often given alone.

There are a number of reasons why the school-to-work movement did not last.[3] Some of the reasons are political: The federal law that fueled the movement was allowed to sunset. But another important reason for the movement's demise was the lack of strong evidence that "school-to-work" reforms produced positive results. There was evidence of sorts, but it was not compelling. For instance, none of the positive results cited in two reviews of the evidence were produced by random assignment evaluations.[4] Of course no one can know whether stronger evidence would have persuaded policymakers to sustain the school-to-work movement. But this is one among many examples of reform movements in education that have come and gone, leaving behind too little enduring knowledge.

Efforts to Improve Education Should Use and Produce Solid Evidence

Efforts to improve education should be guided by evidence of what has worked in the past. And current efforts should continue to collect evidence to inform the future. Few would argue with these assertions.

The more important and difficult question is what evidence to believe. The general criterion we would use is that claims of cause and effect should be clear and subject to a minimum of reasonable doubt.[5] For *practical* purposes, this kind of evidence gives the greatest assurance that a particular strategy will produce the desired effect when applied in new situations.

Among the many perils and pitfalls researchers and evaluators face in trying to make clear causal inferences, we would highlight two: reciprocal causation and selection bias. Both of these are well known, and we have nothing original to say about them. But we find that both are sometimes overlooked in discussions of what is known about effects of high school reform.

Reciprocal causation means that two variables may each be a cause of the other. For instance, suppose a study finds that motivation and grades are positively correlated among a group of students. The explanation could be that stronger motivation has caused some students to study harder and achieve better grades. Or achieving better grades may have caused those students to feel more motivated. Or both could be true. This is one reason for the adage that correlation does not imply causation.

---

[3]In some ways and in some places, the movement continues, but it has certainly lost momentum.

[4]Stern et al., 1995; Urquiola et al., 1997.

[5]For a careful and practical discussion of what cause and effect may mean, see Shadish, Cook, and Campbell, 2002. One useful definition of cause is "an insufficient but non-redundant part of an unnecessary but sufficient condition" (p. 4).

Selection bias occurs when participants in a program or treatment differ from nonparticipants on one or more unmeasured variables that are related to the outcome of interest. For instance, suppose extra instruction were offered to students after the end of the regular school day, and an evaluation compared gains over time for students who did and did not participate. If the participants tend to be students who volunteer because they are more motivated, and if the study does not adequately measure motivation, then the evaluation would overestimate the effect of the program. On the other hand, if teachers specifically recruited the least motivated students and motivation were not measured, the evaluation would underestimate the program's impact. As Heckman's classic paper pointed out, this kind of selection bias is part of the more general class of problems where unmeasured variables are correlated with the outcome and also with one of the measured predictor variables.[6]

Bias can arise not only from the initial selection of participants but also as a result of selective attrition from a program over time. It is usually a fair assumption that students who complete a program differ in unmeasured ways from students who drop out, and that those differences are relevant to what the program was trying to accomplish.

Both problems — reciprocal causation and selection bias — can be avoided by designing an intervention that is relevant to the question being asked and assigning participants at random. For instance, if the question is how much does motivation affect grades, the intervention could be some kind of counseling or experience designed to increase motivation. Randomly assigning participants would ensure that their unmeasured characteristics do not differ much, on average, from nonparticipants if both groups are large.[7]

Random assignment studies in education have limitations, which are also well known. Some questions do not readily lend themselves to experimental manipulation. For instance, if the question is how much do grades affect motivation, it would be difficult to justify random assignment of students' actual grades. Even in situations where an experimental intervention can be designed, it is generally difficult to arrange a uniform nontreatment condition for the control group. Unlike medical research, educational evaluations usually cannot administer a placebo, so the control group receives a mix of "brand X" or "regular school" experiences, and the evaluation becomes a comparison of a fairly well-defined treatment versus a less well-defined set of alternatives. In some educational evaluations, students assigned to the control group even have managed somehow to sneak into the treatment group.

More common is movement in the other direction: Some students who are randomly assigned to a program never actually participate at all. Others begin but leave before completing

---

[6]Heckman, 1979.

[7]How big the samples have to be in order to reduce the average difference to a given level depends on the distribution of the unmeasured variable.

the program. This creates a common dilemma for evaluators: Excluding no-shows or early leavers from the treatment group would defeat the purpose of random assignment, because these students may well differ in unobserved ways from those who do show up and complete the program. But including them dilutes the measured impact of the program. One common procedure to correct for no-shows is to divide the measured impact by the proportion of students assigned to the program who do at least begin it.[8] This procedure makes some plausible assumptions about unmeasured differences between treatment and control groups.[9] Various attempts have also been made to correct for attrition from the treatment group, but these require stronger assumptions about absence of unmeasured differences between early leavers and program completers. If such assumptions were plausible, there would be less need for random assignment in the first place.

Despite these problems with random assignment studies, a well-implemented random assignment design provides the clearest and strongest evidence about cause and effect.[10,11] For this paper, therefore, we tried to find random assignment evaluations of programs that had been implemented in multiple sites. We wanted multiple-site programs because it is most useful to know about strategies that have already been successfully replicated. We also limit consideration here to programs that bear on the institutional design of high schools — arrangements of activities in time and space — not including methods for teaching specific subjects in classrooms.

Some Examples of Solid Evidence

Here we provide brief summaries of three programs that have produced positive impacts for high school students. Each program has been replicated at multiple sites and has been evaluated using random assignment. The impacts we present here are the most conservative estimates reported in the source documents; they are not adjusted for no-shows or degree of participation.[12] Among other outcomes, we focus especially on high school completion, because a

---

[8]Bloom, 1984.

[9]The assumptions are that the program has no effect on students who did not show up, and the probability of being a no-show would have been the same in the control group as in the treatment group. See Myers and Schirm, 1999, p. B-7.

[10]Mosteller, Light, and Sachs, 1996; U.S. Department of Education, 2003.

[11]In situations where random assignment cannot be done, other designs may offer the best evidence possible. Such designs include careful matching of individual participants with nonparticipants who are very similar, use of exogenous instrumental variables as proxies for endogenous differences in educational experience, or interrupted time series analysis of schools or districts before and after a particular intervention. See Shadish, Cook, and Campbell, 2002; Slavin, 2002.

[12]Some of the impacts we report here have been adjusted to take into account measured differences between treatment and control groups. As described in the source documents, these adjustments used either regression or propensity scores. Such adjustments reduce the standard errors of estimated impacts but do not make assumptions about unmeasured variables.

high school diploma becomes more and more important in the labor market as the options available to high school dropouts continue to shrink.[13] We know these three programs are not the only ones that have been evaluated by random assignment, and we do not claim they are the only programs for which there is solid evidence of positive impacts. We present these as exemplars and would be happy to know there are others.[14]

Quantum Opportunity Program (QOP)

Two separate evaluations, both using random assignment, have found that QOP (pronounced "kwop") significantly increased high school completion rates, among other positive outcomes. A community-based organization at each site is responsible for putting in place the QOP model, which combines the following features:

- Each participant has an adult counselor who acts as case manager and advocate. In theory, and often in practice, counselors are accessible to students by telephone or pager 24 hours a day, seven days a week.

- Participants remain in the program whether they change schools, drop out, become incarcerated, or move out of state. The program's motto is, "Once in QOP, always in QOP."

- Educational services include individual assessment and planning, tutoring in high school subjects, and computer-assisted instruction.

- Developmental activities promote life skills and employment readiness, in addition to cultural exposure and recreation.

- Participants perform services that benefit the community.

- Participants are paid about a dollar per hour spent in QOP activities other than recreation or mentoring, and an equal amount is deposited in an accrual account to be used for postsecondary education or training.

---

[13]Levin, 2001.

[14]Our search for solid evidence was greatly facilitated by the excellent compendia of programs compiled by the American Youth Policy Forum, 1997, 1999; Jurich and Estes, 2000; James, Jurich, and Estes, 2001; and the meta-analysis of Comprehensive School Reform model results by Borman et al., 2003.

Participants typically engage in roughly 200 to 300 hours of QOP activity each year. The cost per participant per year is on the order of $5,000 in 2004 dollars. Implementation, participation patterns, and cost vary considerably over time and among sites.[15]

Andrew Hahn and associates at Brandeis evaluated the Ford-funded QOP pilot program in five cities from 1989 to 1993. Hahn summarized the findings.[16] At each site, the evaluators randomly assigned 25 students to QOP and 25 to a control group, from a list of exiting eighth-graders whose families were receiving one or more forms of public assistance. Hahn emphasizes that the evaluation deliberately did *not* require students to apply to QOP, in order to test program operators' ability to recruit low-income students to participate. The Brandeis researchers administered questionnaires in the fall for five years and in the spring of senior year. They also administered tests of academic and functional skills during each year of high school. After the first two years, test scores improved for the QOP group relative to controls. In the fall after scheduled graduation, the survey found these statistically significant differences, among others (p. 247):

	Assigned to QOP	**Control Group**
Percent graduated[a]	63	42
Percent dropped out[b]	23	50
Percent in postsecondary education or training	42	16

NOTES: [a]Hahn (1999) does not indicate whether this includes recipients of General Educational Development (GED) certificates as well as regular diplomas.
          [b]Defined as not having graduated and not currently in school.

Allen Schirm and colleagues at Mathematica evaluated the QOP demonstration funded by Ford in two cities and the U.S. Department of Labor in five cities from 1995 to 2001. Maxfield et al. describe implementation results.[17] Schirm et al. give a detailed analysis of impacts on students.[18] Maxfield et al. summarize both implementation and impacts.[19] Like the Brandeis study, the Mathematica evaluation deliberately did not ask students to apply to QOP. Instead, participants and controls were randomly selected from the population of students in the bottom two-thirds of the grade point average (GPA) distribution among those entering ninth grade for

---

[15]Hahn 1999; Maxfield, Castner, Maralani, and Vencill, 2003; Schirm, Rodriguez-Planas, Maxfield, and Tuttle, 2003.
[16]Hahn, 1999.
[17]Maxfield, Castner, Maralani, and Vencill, 2003.
[18]Schirm, Rodriguez-Planas, Maxfield, and Tuttle, 2003.
[19]Maxfield, Schirm, and Rodriguez-Planas, 2003.

the first time at a high school where the dropout rate was at least 40 percent.[20] However, of the 2,550 students who met these criteria, only 1,069 returned signed consent forms to participate in the evaluation, so there was an element of volition.

In addition to the intake information used to select the sample, Schirm and associates conducted an in-person survey and achievement test in the spring of Year 4 and a telephone survey in Year 5. They also tried to collect transcripts from all high schools that the participants attended. Midway through the year after scheduled high school graduation, the following statistically significant differences emerged:[21]

	Assigned to QOP	Control Group
Percent graduated from regular high school	46	40
Percent with regular diploma or GED certificate or still in high school or a GED program	79	72
Percent in postsecondary education or training[a]	32	26

NOTE: [a]Includes Armed Forces.

The Mathematica evaluation found smaller impacts than the Brandeis study, but it does confirm the earlier findings. These results are important because random assignment studies of other programs to reduce high school dropout rates often have failed to find significant impacts.[22] These two evaluations provide solid evidence that QOP boosts educational attainment by students in populations where high school completion rates are low.

Upward Bound

Created by the federal Higher Education Act in 1965, Upward Bound is a long-established, well-known, and widely distributed program to increase access to college for students whose families have low incomes or whose parents have not attended college. In 1992 the U.S. Department of education began the first large-scale, random assignment evaluation of Upward Bound. The first phase of the study followed most students through high school and some of the older students into postsecondary education.[23]

---

[20] Students deemed by the school to be too disabled to participate in the program were excluded.

[21]Shrirm et al., 2003, Tables V.1, V.3.

[22]See, for example, Dynarski et al., 1998; and Kemple, 2001.

[23]Myers and Schirm, 1999.

Most Upward Bound projects are operated by institutions of higher education, which provide academic counseling, tutoring, and enrichment to participating high school students during the school year and, usually, intensive academic programs on the college campus during the summer. The evaluation classified all Upward Bound projects by type of college sponsor — public or private, two- or four-year — and by urban or rural location, then drew a stratified random sample of projects to represent the program population. Within each project, eligible applicants were randomly assigned to Upward Bound or the control group. In a number of projects, applicants were first classified by characteristics such as race or gender, then randomly assigned within strata. The assignment process occurred over a 14-month period from 1992 to 1994.

In addition to questionnaire data collected on applicants at the time of selection, the first phase of the evaluation conducted telephone surveys and collected school transcripts in 1994-95 and 1996-97. Most students were in Grade 9 or 10 when the study began, and in 1996-97 their high school status was as follows.[24] The difference in the percentage still in high school is statistically significant; the differences in the other two rows are not.

	Assigned to Upward Bound	Control Group
Percent graduated from high school[a]	59	63
Percent still in high school	35	28
Dropped out	6	9

NOTES: [a]Report does not indicate whether this includes recipients of GEDs as well as regular diplomas.

For the sample as a whole, the only other significant impacts as of 1996 were that students assigned to Upward Bound had formed higher expectations regarding their eventual education attainment, and they had completed more high school credits in math and social studies.

The evaluation found more statistically significant impacts for particular subgroups of students (Table V.1).[25] Among students who initially indicated they did not expect to complete a bachelor's degree (21 percent of the study sample), those assigned to Upward Bound were more likely to have graduated from high school by 1996, and they were less likely to have dropped out (Table III.7). Students below the median on an index of academic performance in Grade 9 were also less likely to drop out and more likely to graduate by 1996 if assigned to Upward Bound (Table III.15). Students from low-income families (82 percent of the sample), Hispanics (23 percent), and whites (21 percent) assigned to Upward Bound also were less likely to drop

---

[24]Myers and Schirm 1999, Table III.2.

[25]Table V.1 and the other tables referred to parenthetically in this paragraph can be found in Myers and Schirm (1999).

out, and the Hispanic students were more likely to be still attending high school (Tables III.11, III.13). Boys (29 percent of the sample) were less likely to have dropped out if assigned to Upward Bound (Table III.9). In addition to these impacts on high school status, the evaluation found significant impacts for these same subgroups on educational expectations and the number of credits earned in various high school courses.

A report on the second phase of the evaluation was made available to us as a draft for review.[26] This incorporated results from a survey in 1998-2000. By then, 90 percent of the sample had graduated from high school; 3 percent had obtained GED certificates; and 7 percent had dropped out. There were no significant differences in these outcomes between students who had or had not been assigned to Upward Bound. The only significant difference in high school performance for the sample as a whole was that students assigned to Upward Bound completed more credits in math (Table II.5).[27] Impacts of Upward Bound on high school graduation, drop-out rates, and GED completion also were no longer significant among subgroups defined by low initial educational expectations or weak educational records in Grade 9 (Tables II.6, II.7).

The 1998-2000 survey contained questions about postsecondary education, including names of any schools attended. Evaluators then attempted to obtain respondents' transcripts from those schools. These attempts produced information that either verified the respondent's claim, falsified the claim, or were ambiguous.[28] Using only verified enrollment to calculate enrollment rates may understate true enrollment rates, but using unverified enrollment would overstate them. Myers et al. therefore present both sets of results.[29] For the sample as a whole, the only impact of Upward Bound on postsecondary enrollment was in increase in enrollment at four-year colleges, which was significant in the unverified but not quite significant in the verified data (Tables III.1, III.2).

Among students who had initially indicated they did not expect to obtain bachelor's degrees, the impact on enrollment and credits earned in four-year colleges was significant using both kinds of data (Tables III.3, III.4). Dividing students by academic records in Grade 9, the unverified data showed positive impacts on four-year college enrollment for both high and low achievers, but the verified data showed the impact was significant only for the students who did better in Grade 9 (Tables III.5, III.6). Both verified and unverified data showed a positive impact for Hispanics on enrollment and credits at four-year colleges or other postsecondary

---

[26]Myers et al., 2003.

[27]Table II.5 and the other tables referred to parenthetically in the rest of this section can be found in Myers et al. (2003).

[28]An example of an ambiguous result is an institution responding that it could not release student records without written permission, from which the evaluators could not tell whether the particular student was enrolled or not.

[29]Myers et al., 2003.

schools (Tables III.9, III.10). For both males and females, the unverified data indicated a positive impact on four-year college attendance, but the verified data showed only a positive impact for males on attendance at any postsecondary school (Tables III.11, III.12).

In sum, the first phase of the evaluation indicated that Upward Bound improved high school performance especially for low-income Hispanic and white males who start high school with low educational expectations and weak academic records. However, the follow-up survey three years later, when the entire sample was past high school, found that many of the earlier apparent high school impacts had attenuated or disappeared. Postsecondary impacts were absent or ambiguous for the sample as a whole and for several subgroups. But Upward Bound did increase the rate of four-year college attendance by about 20 percentage points among students who had not expected to earn bachelor's degrees at the time the evaluation began. And among Hispanics, Upward Bound boosted the four-year college-going rate by 12 to 14 percentage points.

Career Academies

The term "career academy" was coined by Stern, Raby, and Dayton to describe a kind of high school program that had originated in Philadelphia in 1969, then spread to California, New York City, and eventually nationwide, encouraged in part by positive results from several quasi-experimental evaluations.[30] There is no authoritative, uniform definition of a career academy, and as the term has become popular, the variation among programs that call themselves career academies has increased.[31] Common themes for career academies are health, business and finance, arts and communications, computers, engineering, law, and government.

In 1993, MDRC began the first random assignment evaluation of career academies.[32] MDRC abstracted three main features to define a career academy:

- School-within-a-school organization, in which academy students at each grade level take a set of classes together and stay with the same small group of teachers from one year to the next.

- A curriculum that includes academic courses meeting college entrance requirements and technical courses — all related to the academy theme.

---

[30]The evidence is summarized by Reller, 1987; and Stern, Raby, and Dayton, 2003.

[31]The State of California provides grants to school districts for "partnership academies," which are defined by statute, but this definition does not apply to the hundreds of academies in California that do not receive state funding. A few other states also have funded such academies. The federal School-to-Work Opportunities Act in 1994 included career academies on a list of seven "promising practices" but did not define them. Building on the MDRC definition, the Career Academy Support Network (http://casn.berkeley.edu) has negotiated a common definition among several networks currently promoting career academies.

[32]Kemple and Rock, 1996.

- Employer partnerships to provide internships and other experiences outside the classroom, related to the academy theme.

The evaluation began with ten sites, but one academy ceased operating. All nine remaining academies are in high schools with large proportions of low-income and minority students. Each was the only career academy in the school.

At the start of the evaluation, the academies recruited more applicants than they could accommodate. Applicants knew that they might not be admitted. MDRC randomly assigned about two-thirds of the applicants to the academy; the others became the control group. In the ten years since the evaluation began, MDRC collected student records, surveyed students during each of their high school years, and conducted follow-up surveys one year and four years after high school.

During the high school years, career academies produced several positive impacts on students' experience and achievement. Compared with the control group, academy students reported receiving more support from teachers and from other students.[33] They were more likely to combine academic and technical courses, to engage in career development activities, and to work in jobs connected to school.[34] As of spring of senior year, academies retained a larger fraction of the students whose initial characteristics made them more likely to drop out.[35] Among students at less risk of dropping out, academies increased participation in technical courses and career development activities without reducing academic course credits.[36]

The first follow-up survey, one year after scheduled graduation, found no significant impacts on students' high school completion, GED acquisition, or participation in postsecondary schooling. It also showed no significant impact on employment or earnings, though students who had been assigned to career academies were working and earning somewhat more than the control group.[37]

The most recent follow-up — about four years after scheduled graduation from high school — found large and significant impacts on employment and earnings and no difference in educational attainment.[38] In the full sample, students who were assigned to career academies earned higher hourly wages, worked more hours per week, had more months of employment, and earned about 10 percent more per month than the control group. All these differences occurred for both males and females, but they were not statistically significant for females. The

---

[33]Kemple, 1997.
[34]Kemple, Poglinco, and Snipes, 1999.
[35]Kemple and Snipes, 2000.
[36]Kemple and Snipes, 2000.
[37]Kemple, 2001.
[38]Kemple, 2003.

MDRC evaluation distinguished between students at high, medium, or low risk of dropping out of high school, as predicted by variables measured before random assignment. Academies had significant positive impacts on average hours worked per week within the 25 percent at high risk, on average hourly wages for the 50 percent at medium risk, and on average monthly earnings for both these groups. Impacts on high school completion or postsecondary education were not significant for the sample as a whole or for any subgroup, but Kemple notes that both the academy and the control group had high rates of high school completion and postsecondary enrollment compared with national data on urban high school students (from the National Education Longitudinal Study, or NELS, of 1988 through 1998).[39]

In sum, the MDRC evaluation found that career academies gave students more personal support, career guidance, technical classes, and school-supervised work experience during high school. Academies also succeeded in retaining more high-risk students through spring of senior year. Eventual impacts on high school graduation or postsecondary education were not significantly positive or negative for the sample as a whole or for any subgroups. But academies had substantial positive impacts on employment and earnings after high school, especially for young men and for students whose initial characteristics indicated high or medium risk of not finishing high school.

A Shared Feature: Accommodating Student Mobility

The three studies described here provide solid evidence that some interventions have produced positive impacts for young people who start high school with poor academic records, low educational expectations, or other challenging circumstances. Although we have focused more on evaluation methods than on program design, we note that all three programs to some extent share a common feature: They can accommodate students who move. QOP explicitly emphasizes trying to stay connected with participants even when they move around, institutionally or geographically: "Once in QOP, always in QOP." Upward Bound also can accommodate some mobility of participants among high schools, because an Upward Bound project typically serves students in several high schools near the college where the project is located. Career academies are less able to keep students who move, because an academy is rooted in its home high school. But some academies do enroll students from other high schools or districts. Accommodating student mobility is important because so many students move in and out of high school or from one school to another, sometimes in the middle of the school year, and students who move more often are less likely to finish high school.

---

[39]Kemple, 2003.

Elusive Inference: Effect of Small Size in High Schools

We turn our attention now to studies that attempt to draw strong causal inferences from evidence not produced by random assignment. To illustrate the difficulty of drawing such inferences, we focus on studies about effects of small size in high schools, a variable that has been given paramount emphasis in current reform strategy. We have not reviewed all the empirical studies on this topic, but we have selected some of the best and most often cited. These studies are informative, and some are ingenious. But they leave considerable room for doubt about the extent to which smaller school size causes better results for students.

The main problem here is the influence of unobserved variables. For example, several frequently cited studies found that smaller high schools have lower dropout rates.[40] Each study compared high schools in a state or national sample at one point in time. Some of the smaller high schools would be located in smaller, close-knit suburban or rural communities — the kind of place where teachers and administrators send their own children to the school where they work. Students who cut classes are more likely to be caught if they live in a community where more people know one another, so cutting classes would be less likely and would less often lead to dropping out of school entirely. In big cities, more of the small high schools would be magnets or other schools of choice. In these situations as well, stronger social cohesion and shared values among parents and teachers could account for the lower dropout rates. The density of personal connections and strength of shared expectations among parents and school staff are unmeasured variables in these studies. Socioeconomic variables used as statistical controls do not capture these differences. The association between smaller school size and lower dropout rates, therefore, could be due at least partly to smaller high schools' occurring in particular kinds of circumstances that account for the better results.

Unmeasured variables also may influence the selection of certain kinds of students into particular small schools, or into smaller subschools within large high schools. Various studies have found that students in smaller schools are relatively less alienated, more engaged, and more likely to pass courses and earn credits toward graduation.[41] Studies also have found better student performance in smaller learning communities (SLCs) within large urban high schools.[42,43]

However, these results may be largely attributable to small schools' or SLCs' enrolling students whose unmeasured, preexisting characteristics would have made them more likely to perform better in any situation. In metropolitan areas, small schools are often magnets, alternative schools, or other schools of choice. Similarly, SLCs within larger high schools usually en-

---

[40]Fetler, 1989; Franklin and Crone, 1992; Howley and Bickel, 1999; Pittman and Haughwout, 1987.
[41]See reviews by Cotton,1996; Gladden, 1998; Raywid, 1995.
[42]Stern (2003) reviews these studies in more detail.
[43]McMullan et al., 1994; Oxley, 1990; Wasley et al., 2000.

roll students who choose to be there. Students who are more motivated or better organized, or whose parents are more concerned about their schooling, may be more likely to exercise choice in the first place. Schools and SLCs naturally seek to enroll and retain students with these kinds of qualities. These characteristics of students and families, not measured by researchers, could account in part for the students' better performance. The ongoing process of mutual selection may result in small schools' or SLCs' enrolling more students whose unmeasured, preexisting characteristics would make them more likely to succeed anywhere. One indication of this dynamic is the finding by Wasley et al. that a lower dropout rate among SLC students occurred in high schools where only some students were in SLCs but not in high schools where all students were in SLCs.[44] In instances where converting an entire school to SLCs has led to better outcomes, it may not be clear whether some low-performing or misbehaving students who would have attended the school before the transformation did not enroll there after the change.[45]

One way to avoid selection bias in testing whether small school size causes better student performance would be to use random assignment. Students could be randomly assigned to large schools, small schools, or SLCs. We have not yet found such studies. High school programs in the random assignment evaluation of the School Dropout Demonstration Assistance Program all had small enrollments.[46] None of these programs increased the proportion of students earning regular high school diplomas, but the focus of this evaluation was not small size per se.

Another approach would be to randomly assign entire high school attendance zones or school districts to enroll in large or small schools. We have not yet found such a study. Gottfredson did observe what happened in five high schools where major enrollment changes suddenly occurred as a result of district reorganization.[47] In two high schools that became bigger, there was no change in reported drug use or delinquency; teachers' expressed feeling of safety decreased in one school; and students' reports of victimization by other students increased in one school. In three high schools that became smaller, reports of drug use and delinquency increased in two schools; teachers' feeling of safety improved in one school; and students' reports of victimization increased in one school. These results probably reflect some changes in student population as well as change in school size, but they do not indicate that the size change was decisive.

Current strategies to improve high schools seldom rely on smaller size alone. Lee and Smith argue that small size itself is not a direct cause of better student performance but that

[44]Wasley et al., 2000.
[45]See, for example, McPartland et al., 1998.
[46]Dynarski et al., 1998.
[47]Gottfredson, 1985.

"smaller school size is a facilitating factor for creating organizational features of schools that we have shown to be important determinants of learning."[48] Those organizational factors include teachers' sense of collective responsibility for learning, students' taking more math and science courses, and use of more authentic instructional practices (Table 6.3). These findings are derived from an elegant statistical analysis of NELS data, using hierarchical models to distinguish between the connection of school characteristics to average achievement (excellence) and their connection to the within-school correlation of achievement with socioeconomic status (equity).

Lee and Smith's analysis of high schools is theoretically strong and empirically sophisticated. Nevertheless, it leaves open several questions about the effects of school size. Lee and Smith do not present evidence that smaller school size is associated with teachers' sense of collective responsibility for learning, students' taking more math and science courses, or use of more authentic instructional practices.[49] Even if these characteristics are more apt to be present in smaller schools, the observed association between these school features and student learning could be attributed to reciprocal causation. For instance, Lee and Smith measure teachers' sense of collective responsibility by their responses to 12 survey items including, "I can get through to the most difficult student," "Teachers make a difference in students' lives," and (with reverse scoring) "Students are incapable of learning the material" (p. 190). Teachers may be more inclined to give positive answers to these and the other items as a *result* of being in a school where students are more successful. Likewise, students may take more math and science courses, and may be exposed to more challenging instructional methods, *because* they are successful learners. So it is not clear to what extent these school characteristics are the cause or effect of student learning.

Except in one chapter, Lee and Smith's statistical models include high school enrollment as a single number among other school characteristics in a linear combination of predictors.[50] But in a separate chapter focusing on size itself, Lee and Smith divide schools into eight categories, by enrollment, and find that students in schools with enrollments of 600 to 900 had the biggest average gains in achievement, compared with larger or smaller schools.[51] This result raises additional questions. Are the school characteristics they found to be associated with student learning also most prevalent in this same size category, compared with schools that are larger or smaller? Are there other, unmeasured characteristics of schools that may be concentrated in this size range? For instance, community characteristics may be different in very small rural schools or very large urban schools, compared with medium-sized schools in suburbs or small

---

[48]See Lee and Smith, 2001.

[49]See Lee and Smith, 2001.

[50]Lee and Smith, 2001.

[51]The distribution of achievement was least associated with socioeconomic status in all size categories below 600, compared with bigger schools.

towns. Or a larger proportion of schools in the 600 to 900 range may be magnets or other schools of choice.

A study that viewed size as one factor among others, and also gave careful attention to student selection, is the account by Darling-Hammond, Ancess, and Ort of changes at Julia Richman High School in New York City.[52] They point out that small size is not a sufficient condition for improvement: "Not all small schools are successful" (p. 642). They describe the transformation of Julia Richman from a large high school into a set of small, autonomous schools sharing the same site. The new, small schools built strong relationships between and among students and faculty by reducing the pupil load for each teacher and creating new advising structures; developed more coherent curriculum; engaged students in active learning; used portfolios and exhibitions to assess students' work; and provided time for teachers to collaborate.

Darling-Hammond, Ancess, and Ort paid careful attention to possible selection bias. In the new schools' first year, the student body comprised mostly students from the Julia Richman attendance zone "who had not applied elsewhere or had been rejected by their chosen school" (p. 645).[53] Seventy percent were eligible for subsidized lunch, compared with 32 percent of students at Julia Richman in the previous year. Some selective attrition occurred in the first couple of years, as many students who "had not proactively chosen the schools" moved out (p. 648). But analysis of students who entered the new schools as ninth-graders in 1994, excluding transfers in or out, found a four-year graduation rate of 73.3 percent, "significantly higher than the comparable New York City rate of 49.7 percent for the same cohort" (p. 649). Six-year graduation rates were also higher. Even though some transfers out of the comparison schools would graduate from a school other than the one they entered in ninth grade, these results suggest that the new schools at the Julia Richman site had stronger than average holding power. In addition, eleventh-graders (presumably including those who transferred in) at the new small schools outperformed students in similar schools on New York State Regents examinations for reading and writing, though not for math. Among graduates from the new schools, college-going rates were 86 percent in 1997 and 91 percent in 1998.

The attention given to selection and attrition makes this a more persuasive study.[54] How much of the observed effect is attributable to the new schools' small size remains unclear.

Selection and Choice

The likelihood of selection bias pervades much of the existing research on effects of small high schools and small learning communities (SLCs) within large high schools. If the ap-

---

[52]Darling-Hammond, Ancess, and Ort, 2002.
[53]Darling-Hammond, Ancess, and Ort, 2002.
[54]Darling-Hammond, Ancess, and Ort, 2002.

parent positive results of small size are largely due to selection of students with positive un-measured traits such as motivation, then transforming all large high schools into smaller ones would not accomplish much. In addition, teachers in new small schools or SLCs also may be self-selected. If these teachers possess more motivation, commitment, energy, creativity, or other positive traits, then positive results from these small settings may not generalize to the system as a whole. There is a danger that current attempts to downsize all high schools may be based on a fallacy of composition — a mistaken hope that what is observed in specific cases can be generalized to the whole high school population.

Although self-selection of students and teachers makes it more difficult to draw clear causal inferences, self-selection could be a good thing in a programmatic sense. It is possible that particular high schools or SLCs are good for students who choose them but not for other students. If that were the case, the best arrangement might be to let students choose from an array of large and small schools or SLCs. Some large districts are already doing that.

Whether expanding school choice improves outcomes for students is itself a vigorously contested empirical question. Studies to date have focused mainly on elementary schools, but Cullen, Jacob, and Levitt have studied effects of high school choice in Chicago.[55] To control for the possibility that students applying to a particular school might share certain unobserved characteristics, they focused on 19 high schools that used random lotteries to select students. They found that students who won a lottery at the time they entered ninth grade did *not* perform better academically in Grade 9 or 10, compared with students who did not win in the same lottery. As economists, they viewed these findings as "surprising" (p. 4). In another paper (forthcoming), the same authors used proximity to different kinds of high schools as exogenous instruments to estimate the effects of choosing to enroll in one of 12 high-achieving schools, 10 career academies, or 39 other schools. They focused on whether students successfully completed high school, and they found positive effects only for the career academies.[56]

Studies Using Data for Whole School Districts

One way to reduce possible selection bias is to study the whole student population in a big school district. For example, if a school district increased the number of small high schools or SLCs, evidence on districtwide trends in student performance could reveal the extent of gains for students choosing these options, as well as any possible negative trends among the students

---

[55]Cullen, Jacob, and Levitt, 2003.
[56]Career academies in Chicago are different from the model described above and evaluated by MDRC. Chicago's career academies are full-sized high schools that emphasize career and technical education.

left behind.[57] McMullan, Sipe, and Wolf did this kind of analysis in Philadelphia, where the district, encouraged by the Pew Charitable Trust, greatly expanded the number of high school SLCs (called "charters") from 1988-89 to 1993-94.[58] The proportion of high school students enrolled in SLCs rose steadily over this period, but most districtwide indicators of academic performance, after some initial gains, leveled off or went back down. The authors suggest that the gains due to SLCs might have been offset in the later years by changes in district policy that moved more over-age middle-school students into high schools and also cut summer school.

Another study of districtwide effects was the evaluation by Bohrnstedt et al. of Equity 2000, a program by the College Board to increase math course-taking, college preparation, and college enrollment among low-income Hispanic and African-American students.[59] Results are reported for six urban districts that enacted policies for all students to take first-year algebra by Grade 9 and geometry by Grade 10, and the districts provided various kinds of support for teachers to make this happen. Course-taking and other outcomes were measured by surveys given to all graduating seniors in three successive cohorts.[60] Results show that larger proportions of students in the later cohorts took algebra by Grade 9 and geometry by Grade 10. Increases in geometry course enrollment by Grade 10 were greater for Hispanics and African-Americans than for Asians or whites. However, there were no apparent gains in the proportions of students taking advanced math courses or college entrance examinations.

Snipes, Doolittle, and Herlihy used districtwide data in a study of successful urban districts.[61] Several urban districts were chosen from different parts of the country, based on evidence that student achievement had improved for at least three years and that differences in average achievement between white and minority students had narrowed. As in the "effective schools" studies of the 1970s and 1980s, the purpose here was not to test the impact of an intervention that was defined ex ante but, instead, to try to identify practices that might account for these districts' apparent success. To strengthen inferences about which practices mattered, the MDRC study also visited two comparison districts that were similar in some ways but had not improved student achievement. We mention this study as an example of the districtwide method, but — unfortunately, for our purposes — the districts that were studied had not experienced gains in student achievement at the high school level (pp. 106-109, 138-141).

---

[57]Selection bias may still occur if there is substantial change in the numbers or kinds of students who enroll in private schools or other districts.

[58]McMullan, Sipe, and Wolf, 1994.

[59]Bohrnstedt et al., 1999.

[60]Response rates by cohort were 52 percent, 61 percent, and 64 percent (p. 14).

[61]Snipes, Doolittle, and Herlihy, 2002.

Studies of Comprehensive High School Reform Models

A considerable amount of recent and ongoing effort has been focused on evaluating federally identified "comprehensive school reform" (CSR) models, but strong evidence is not yet available on CSR models at the high school level. An extensive meta-analysis by Borman et al. (2003) summarizes the effects on student achievement of 29 widely implemented CSR models. Only two models were designed specifically for high schools, Grades 9-12: High Schools That Work (HSTW) and Talent Development High Schools (TDHS). In contrast, there were four CSR models for elementary Grades K-5, seven models for Grades K-8, and 16 CSR models for Grades K-12. Some of the K-12 models have been studied in high schools, but the meta-analysis combined into one category all studies that included any students in Grades 6-12.

Of the 232 studies that met the inclusion criteria for the CSR meta-analysis, 45 reported measures of student achievement from HSTW and one from TDHS.[62] All these studies were sponsored by the models' developers, except one study of HSTW. None of these studies used random assignment. The HSTW studies also rely mainly on senior-year data from successive cohorts of students who completed defined sequences of academic and vocational courses. Changes in HSTW results over time may reflect changes in the composition of the students selected.

Evaluations of CSR models are continuing, however. It is possible that one or more models may yet produce solid evidence of effects for high school students.

Recommendations

Our purpose here was to illustrate the application of strict scrutiny to claims of cause and effect in studies of programs or strategies for high school students. We described three examples of multisite evaluations that produced solid evidence of positive impacts. We hope there are other examples already published or forthcoming. Since our search was not exhaustive, we do not claim to provide a comprehensive review of everything known to be effective for high school students. Nevertheless, we will offer two recommendations.

First, increase investment in long-term evaluations using random assignment. The three evaluations we describe each took about a decade to produce clear findings. Given the severity of problems in American high schools, attempts to make improvement must go forward. But more of these attempts should be accompanied by random assignment evaluations. In some situations, such as initiatives to expand choice among schools or small learning communities within schools, the use of lotteries to select students provides a natural opportunity for this kind of evaluation. Even when it is not built into the program, random assignment of students, class-

---

[62]McPartland et al., 1998.

rooms, schools, or entire districts should be done more often, and more resources should be spent on data collection and analysis. Assigning people to control groups does bar them from interventions that may be beneficial, and spending more on evaluation may take money from program operation. But these harms may be less than the possible damage caused by promoting massive changes without good evidence that they are producing desired results.

Second, on a more positive note, the random assignment evaluations of QOP, Upward Bound, and career academies have produced solid evidence on which to build. In addition to justifying more replication of these programs in their current forms, the results may point the way to further development, evolution, or hybridization of these initiatives.[63] We noted, in particular, that the three programs — especially QOP — all accommodate students who move. This is important because students who change schools more often are less likely to finish high school successfully. Many current high school reforms are attempting to build small learning *communities,* intended to nurture sustained interpersonal relationships, from which students can benefit only if they stay there for some period of time. QOP has shown that it is possible to form a relationship that continues for several years between a high school student and a caring adult, even when the student does not remain in one place.

References

American Youth Policy Forum. 1997. *Some Things DO Make a Difference for Youth: A Compendium of Evaluations of Youth Programs and Practices*. Washington, DC: American Youth Policy Forum.

American Youth Policy Forum. 1999. *More Things That DO Make a Difference for Youth: A Compendium of Evaluations of Youth Programs and Practices.* Vol. II. Washington, DC: American Youth Policy Forum.

Bloom, H. 1984. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review* 8: 225-246.

Bohrnstedt, G., P. Jakwerth, C. Rodriguez, and S. Quiñones. 1999. *The Senior Survey Analysis of Cohorts 1, 2, and 3.* Report No. 87 (September). Palo Alto, CA: American Institutes for Research, John. C. Flanagan Research Center.

Borman, G. D., G. M. Hewes, L. T. Overman, and S. Brown. 2003. "Comprehensive School Reform and Achievement: A Meta-Analysis." *Review of Educational Research* 73 (2) (Summer): 125-230.

Cotton, K. 1996. *Affective and Social Benefits of Small-Scale Schooling.* ERIC Digest No. RC 96-5. Charleston, WV: Clearinghouse on Rural Education and Small Schools. ERIC Document Reproduction Service No. ED 401 088.

---

[63] Any new versions or hybrids should be rigorously evaluated, of course!

Cullen, J. B., B. A. Jacob, and S. D. Levitt. 2003. "The Effect of School Choice on Student Outcomes: Evidence from Randomized Lotteries." Working Paper 10113 (November). Cambridge, MA: National Bureau of Economic Research.

Cullen, J. B., B. A. Jacob, and S. D. Levitt. Forthcoming. "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools." *Journal of Public Economics.*

Darling-Hammond, L., J. Ancess, and S. W. Ort. 2002. "Reinventing High School: Outcomes of the Coalition Campus Schools Project." *American Educational Research Journal* 39 (3): 639-673.

Dynarski, M., P. Gleason, A. Rangarajan, and R. Wood. 1998. *Impacts of Dropout Prevention Programs: Final Report.* MPR Reference No.: 8014 (June). A research report from the School Dropout Demonstration Assistance Program Evaluation, submitted to the U.S. Department of Education Planning and Evaluation Service. Princeton, NJ: Mathematica Policy Research, Inc.

Elliott, M. N., L. M. Hanser, and C. L. Gilroy. 2000. *Evidence of Positive Student Outcomes in JROTC Career Academies.* Santa Monica, CA: RAND Corporation.

Fetler, M. 1989. "School Dropout Rates, Academic Performance, Size, and Poverty: Correlates of Educational Reform." *Educational Evaluation and Policy Analysis* 11: 109-116.

Franklin, B. J., and L. J. Crone. 1992. *School Accountability: Predictors and Indicators of Louisiana School Effectiveness.* Paper presented at the meeting of the Mid-South Educational Research Association, Knoxville, TN, November.

Gladden, R. 1998. "The Small School Movement: A Review of the Literature." Pages 113-137 in M. Fine and J. I. Somerville (Eds.), *Small Schools, Big Imaginations: A Creative Look at Urban Public Schools.* Chicago: Cross City Campaign for Urban School Reform.

Gottfredson, D. 1985. *School Size and School Disorder.* Baltimore: Johns Hopkins University, Center for Social Organization of Schools.

Hahn, A. 1999. "Extending the Time of Learning." In D. J. Besharov (Ed.), *America's Disconnected Youth: Toward a Preventative Strategy.* Washington, DC: Child Welfare League of America Press.

Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1) (January): 153-161.

Howley, C. B., and R. Bickel. 1999. *The Matthew Project: National Report.* Albany, OH: Ohio University.

James, D. W., S. Jurich, and S. Estes. 2001. "Raising Minority Academic Achievement: A Compendium of Education Programs and Practices." Washington, DC: American Youth Policy Forum.

Jurich, S., and S. Estes. 2000. "Raising Academic Achievement: A Study of 20 Successful Programs." Washington, DC: American Youth Policy Forum.

Kemple, J. J. 1997. *Career Academies: Communities of Support for Students and Teachers: Emerging Findings from a 10-Site Evaluation.* New York: MDRC.

Kemple, J. J. 2001. *Career Academies: Impacts on Students' Initial Transitions to Post-Secondary Education and Employment.* New York: MDRC.

Kemple, J. J. 2004. *Career Academies: Impacts on Labor Market Outcomes and Educational Attainment.* New York: MDRC.

Kemple, J. J., S. M. Poglinco, and J. C. Snipes. 1999. *Career Academies: Building Career Awareness and Work-Based Learning Activities Through Employer Partnerships.* New York: MDRC.

Kemple, J. J., and J. L. Rock. 1996. *Career Academies: Early Implementation Lessons from a 10-Site Evaluation.* New York: MDRC.

Kemple, J. J., and J. C. Snipes. 2000. *Career Academies: Impacts on Students' Engagement and Performance in High School.* New York: MDRC.

LaPoint, V., W. Jordan, J. M. McPartland, and D. P. Towns. 1996. *The Talent Development High School: Essential Components.* Baltimore: Johns Hopkins University and Howard University, Center for Research on the Education of Students Placed At Risk.

Lee, V. E., and J. B. Smith. 2001. *Restructuring High Schools for Equity and Excellence: What Works.* New York: Teachers College Press.

Levin, H. M. 2001. "High-Stakes Testing and Economic Productivity." Pages 39-50 in G. Orfield and C. Edley (Eds.), *Raising Standards or Raising Barriers? Inequality and High-Stakes Testing in Public Education.* New York: Century Foundation.

Maxfield, M., V. Maralani, and M. Vencill. 2003. *The Quantum Opportunity Program Demonstration: Implementation Findings.* MPR Reference No. 8279-080 (August). Washington, DC: Mathematica Policy Research, Inc.

Maxfield, M., A. Schirm, and N. Rodriguez-Planas. 2003. *The Quantum Opportunity Program Demonstration: Implementation and Short-Term Impacts*. MPR Reference No. 8279-093 (August). Washington, DC: Mathematica Policy Research, Inc.

McMullan, B. J., C. L. Sipe, and W. C. Wolf. 1994. *Charters and Student Achievement: Early Evidence from School Restructuring in Philadelphia.* Philadelphia: Center for Assessment and Policy Development.

McPartland, J. M., R. Balfanz, W. Jordan, and N. Legters. 1998. "Improving Climate and Achievement in a Troubled Urban High School Through the Talent Development Model. *Journal of Education for Students Placed at Risk* 3: 337-361.

Mosteller, F., R. J. Light, and J. A. Sachs. 1996. "Sustained Inquiry in Education: Lessons in Skill Grouping and Class Size." *Harvard Education Review* 66: 707-842.

Myers, D., R. Olsen, N. Seftor, J. Young, and C. Tuttle. 2003 *The Impacts of Regular Upward Bound: Results from the Third Follow-Up Data Collection*. Draft Report. Washington, DC: Mathematica Policy Research, Inc.

Myers, D., and A. Schirm. 1999. *The Impacts of Upward Bound: Final Report for Phase I of the National Evaluation.* MPR Reference No.: 8046-515 (April). Submitted to the U.S. Department

of Education Planning and Evaluation Service. Washington, DC: Mathematica Policy Research, Inc

Oxley, D. 1990. *An Analysis of House Systems in New York City Neighborhood High Schools.* Philadelphia: Temple University Center for Research in Human Development and Education.

Pittman, R. B., and P. Haughwout. 1987. "Influence of School Size on Dropout Rate." *Educational Evaluation and Policy Analysis* 9: 337-343.

Raywid, M. A. 1995. *The Subschools/Small Schools Movement — Taking Stock.* Madison, WI: Center on Organization and Restructuring of Schools.

Reller, D. J. 1987. *A Longitudinal Study of the Graduates of the Peninsula Academies, Final Report.* Palo Alto, CA: American Institutes for Research in the Behavioral Sciences.

Rodriguez, C., N. Khattri, and Mei Han. 1997. *The Equity 2000 Evaluation, a Summative Report: Impact and Implementation.* Report No. 86 (June). Washington, DC: Pelavin Research Center.

Schirm, A., N. Rodriguez-Planas, M. Maxfield, and C. Tuttle. 2003. *The Quantum Opportunity Program Demonstration: Short-Term Impacts.* MPR Reference No.: 8279-093 (August). Washington, DC: Mathematica Policy Research, Inc.

Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin.

Slavin, R. E. 2002. "Evidence-Based Education Policies: Transforming Educational Practice and Research." *Educational Researcher* 31 (7): 15-21.

Snipes, J., F. Doolittle, and C. Herlihy. 2002. *Foundations for Success: Case Studies of How Urban School Systems Improve Student Achievement.* New York: MDRC.

Southern Regional Education Board. 1995. *Charting the Progress of Education: Annual Report* (June). Atlanta, GA: Southern Regional Education Board.

Southern Regional Education Board. 1997. *1997 Outstanding Practices.* Atlanta, GA: Southern Regional Education Board.

Stern, D. 2003. "Career Academies and High School Reform Before, During, and After the School-to-Work Movement." Pages 239-262 in W. J. Stull and N. M. Sanders (Eds.), *The School-to-Work Movement, Origins and Destinations.* Westport, CT: Praeger.

Stern, D., N. Finkelstein, J. R. Stone III, J. Latting, J., and C. Dornsife. 1995. *School to Work: Research on Programs in the United States.* Washington and London: Taylor and Francis, Falmer Press.

Stern, D., M. Raby, and C. Dayton. 1992. *Career Academies: Partnerships for Reconstructing American High Schools.* San Francisco: Jossey-Bass.

Urquiola, M., D. Stern, I. Horn, C. Dornsife, B. Chi, L. Williams, D. Merritt, K. Hughes, and T. Bailey. 1997. *School to Work, College and Career: A Review of Policy, Practice, and Results 1993-1997.* Berkeley, CA: National Center for Research in Vocational Education.

U.S. Department of Education. 2003. *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Districtwide Strategies of
High School Reform

The Case of Chicago

G. Alfred Hess, Jr.

Introduction

Reform initiatives that encompass all schools in a district represent very different opportunities and complexities than do efforts that focus on individual schools or groups of schools. They also present different challenges for researchers attempting to understand their successes and shortcomings. This paper will attempt to address some of the differences associated with districtwide reform efforts, using Chicago as a case study.

Setting the Scene

The Chicago Public Schools (CPS) have been engaged in serious reform efforts for 15 years, beginning with the local school governance reforms legislatively mandated in 1988 and the more recent, centrally directed high stakes accountability reforms initiated in 1996. The first set of reforms focused on local decision-making, encouraging many different kinds of efforts to improve student achievement, including schools that chose to reject all improvement efforts.

The second set of reforms applied more evenly to all schools, particularly at the high school level, where a new design for high schools was adopted by the Board of Education. During the 1980s, CPS was a typical urban school district, serving a predominantly minority student enrollment. In 1980, about 18 percent of the students were white; by 1990, white enrollment had fallen to about 11 percent. In 1978, a state-mandated reassignment of teachers created faculties at each school that had no more than 60 percent of any race, meaning that in most predominantly African-American schools, 60 percent of the faculty were African-American; while in predominantly Hispanic, white, or mixed schools, the faculty were 60 percent white. Most faculty lunchrooms were marked by racially distinct teacher groupings. In 1980, the school district could not meet its payroll and was put into a kind of financial receivership that resulted in the elimination of 8,000 positions (one-sixth of the total); most of the positions that were eliminated were teachers or teacher's aides. Part of the remedy for the fiscal crisis was the appointment of an entirely new Board of Education, which for the first time included a majority of African-American and Hispanic members. Shortly thereafter, the board hired the city's first African-American superintendent.

Two nonprofit research and advocacy agencies issued a series of widely publicized reports on the ineffectiveness of the district's schools, including the fact that nearly half of the district's entering ninth-graders never graduated from high school. Reading and math achievement were generally a year to a year and a half below the national norms. Harold Washington — elected in 1983 as the city's first African-American mayor — convened a blue-ribbon task force in 1986 to come up with ways to improve the school system. Faced with administrative resistance to any serious changes in the school system, and a teachers' strike that delayed the opening of school for more than a month in the fall of 1987, the mayor launched an expanded

reform effort that brought together business leaders, community activists, and educational reform advocates in a movement that resulted, after Washington's death, in the 1988 school-based governance legislation.[1]

The Chicago School Reform Act of 1988 (P.A. 85-1418) had three major components. The first was a set of 10 goals, the main thrust of which was that Chicago students should perform in reading and math at levels comparable to other students across the country. This goal came to be known as "meeting national norms" and eventually would be measured in terms of 50 percent of students reading and doing math at or above the national medians. The reform act applied this goal not only to the district as a whole but also to each individual school in the district. The second component of the act required that all schools be provided an equal base of educational programs, on top of which state poverty funds would be distributed on the basis of the enrollment of students qualifying for a free or reduced-price lunch. These poverty-generated funds would be used at the discretion of local school leaders. When these new discretionary funds were fully phased in over five years, the average elementary school received about $500,000 annually, and the average high school received $750,000. The third component of the reform act created Local School Councils (LSCs) at each school, composed of six parents, two community representatives, two teachers, and the principal. The LSC would have the power to decide how the school's budget would be expended, including the discretionary money provided by the act; what the school's improvement plan would be; and who would be the principal. The power to hire and fire the principal resulted, over the next five years, in an 80 percent turnover in school leadership, and many more African-Americans and Hispanics assumed the principalship.

During the early 1990s, three main streams emerged among elementary schools across the district.[2] About a third of the schools made significant efforts to radically reorganize their schools for improvement. About a third of the schools utilized their new discretionary funds to add many different programs; these schools were characterized as "Christmas tree" schools, without a central focus. About a third of the schools made no significant efforts at change or improvement. Reading and math achievement inched upwards in the elementary schools but plunged in the high schools (see Figures 1 and 2). Some individual schools showed significant increases in achievement, but most made little progress.

In 1995, in response to a new round of fiscal crises, and with the recognition that elected LSCs did not provide enough leverage to force academic improvement on all schools,

---

[1] For a fuller account of the shortcomings of the school district, the movement to reform the schools, and the 1988 reform act, see Hess (1991).
[2] Bryk et al., 1993.

Figure 1

Elementary School Achievement
Chicago Public Schools

Percent At/Above Norms — Reading, Math

Math: 27.1 (1990), 26.6 (1991), 26.3 (1992), 30.1 (1993), 28.0 (1994), 29.8 (1995), 31.2 (1996), 35.9 (1997), 39.6 (1998), 43.6 (1999), 46.7 (2000)

Reading: 23.5 (1990), 21.8 (1991), 22.4 (1992), 26.8 (1993), 26.5 (1994), 26.5 (1995), 29.1 (1996), 30.3 (1997), 34.7 (1998), 36.0 (1999), 36.4 (2000)

3rd - 8th Grades

Figure 2

High School Achievement
Chicago Public Schools

Percent At/Above Norms — Reading, Math

Reading: 30.6 (1990), 27.6 (1991), 23.7 (1992), 25.5 (1993), 22.2 (1994), 23.6 (1995), 20.5 (1996), 24.6 (1997), 28.6 (1998), 32.5 (1999), 35.3 (2000)

Math: 21.6 (1990), 21.6 (1991), 21.3 (1992), 25.4 (1993), 20.7 (1994), 25.3 (1995), 21.7 (1996), 30.4 (1997), 31.1 (1998), 41.9 (1999), 45.0 (2000)
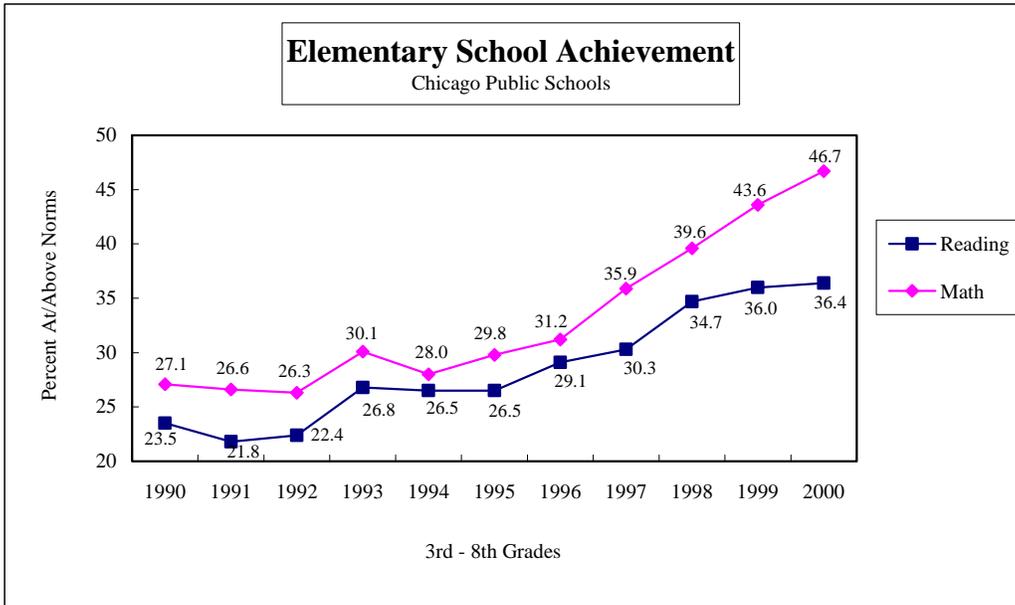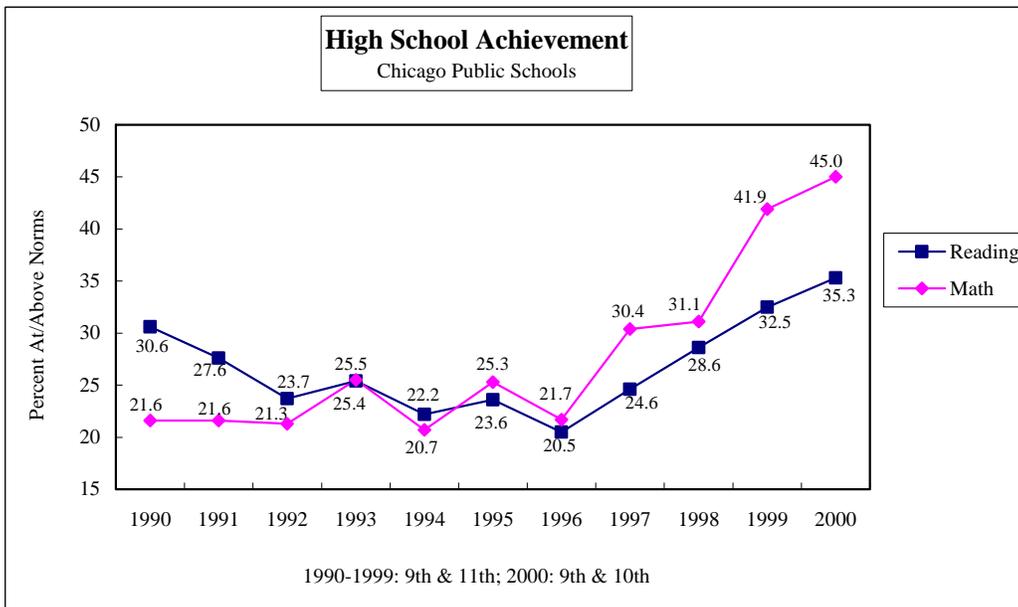
1990-1999: 9th & 11th; 2000: 9th & 10th

the state legislature amended the Chicago School Reform Act to vest control of the central administration in the mayor's office and to enact a set of measures by which the newly constituted Reform Board of Trustees (a smaller, more corporate style board of education) could hold individual schools accountable for student achievement. For the first time, there would be serious consequences for schools in Chicago if student achievement did not measure up. The mayor, Richard M. Daley, appointed the city budget director, Paul Vallas, as chief executive officer of the school system and appointed his just-departed chief of staff, Gery Chico, as president of the Reform Board. By 1996, Vallas had resolved the system's financial crisis and had refinanced the district's long-term debt, freeing up hundreds of millions of dollars. He had also instituted a high-stakes accountability system that placed more than 100 schools on probation and had announced a plan to "end social promotion," installing promotion gates at third, sixth, and eighth grades that would retain up to 10,000 low-performing students per year in their prior grade.

Meanwhile, high school reading achievement had plummeted during the early 1990s. In 1996, Vallas convened a task force of more than 200 stakeholders, both from within the school system and from across the city, to formulate a plan to turn around the city's failing high schools. The resulting plan, somewhat amended, was adopted by the Reform Board in March 1997 as a new Design for High Schools.[3]

The Design for High Schools incorporated new, higher graduation requirements; it mandated that all high school students take a full schedule of credit-bearing courses; it required high schools to create advisories and Freshmen Academies in an effort to overcome anonymity; and it included the district's school accountability provisions of probation and possible reconstitution. The two themes of the new high school design were increasing the "press" for academic improvement and enhancing personalism as a way to support students' being asked to meet higher standards. Some 38 of the city's 75 high schools had already been placed on probation for having less than 15 percent of their students at or above the national norm in reading or math. Not only did probation bring a reputational sanction, but schools also lost the ability to determine who the principal would be. Principals could be dismissed and replaced by central-office appointees, who served at the pleasure of the Reform Board. But probation also brought additional resources. Probation managers were appointed to provide mentoring to principals; funds were provided (about $100,000 per year) to hire external partners to provide assistance to schools in their efforts to improve. The percentage of students meeting national norms in reading and math became the primary criterion for judging school improvement. In the summer of 1997, additional teeth were added to school accountability when seven high schools were reconstituted, with all positions at the schools being declared vacant. By the time school opened in

---

[3]Chicago Public Schools, 1997.

the fall, more than 30 percent of the teachers in these schools had been replaced, as had five of the seven principals.

The Research Effort in Chicago

In addition to significantly impacting the city's public schools, the Chicago School Reform Act engendered an important expansion of research on urban schools and the efforts to improve them and to reform the system in which they were located. As noted above, the primary research that focused on the Chicago Public Schools prior to 1988 had been undertaken by two nonprofit research and advocacy organizations, Designs for Change and the Chicago Panel on School Policy. Between the two organizations, about a dozen major research reports were released during the 1980s. (Appendix B provides a bibliography of research on the Chicago Public Schools.) That research focused on the misuse of special education, the system's finances and resource utilization, school effectiveness, and dropping out. The effective schools research provided a theoretical framework behind much of the work of these two organizations and provided a foundation for the 1988 Chicago School Reform Act.[4] Most schools of education in the Chicago region had paid little attention to CPS and conducted little, if any, research in the system's schools. After the adoption of the reform act, the Chicago Panel on School Policy received major foundation support to conduct a mixed-methods evaluation of the school-based management reforms and their impact on school resources and student achievement.

The reform act mandated that the school district mount an extensive research effort to monitor the effects of reform implementation at the same time that it reduced the size of the central administration to free up discretionary funds in the schools. The CPS director of research asked Professor Tony Bryk at the University of Chicago to convene a meeting of university and nonprofit researchers to develop a collaborative effort to monitor the reform's implementation. From these meetings, the Consortium for Chicago School Research was born, led by Bryk and a set of directors drawn from the major area universities, nonprofit research organizations, and the leadership of the school system. The consortium launched a set of biannual surveys of principals, teachers, and students. It also undertook a number of specific research projects, engaging scholars from across the nation to examine aspects of the reform efforts in Chicago. In the mid-1990s it became the Chicago partner in a nationwide evaluation of the Annenberg Challenge. The consortium became the largest producer of research on the Chicago reforms. Its relationship with the school system waxed and waned under different system administrations. For two years, Bryk led the district's research efforts while on partial released time from the university. The consortium's relationships with the Vallas administration were more strained.

---

[4]See Brookover and Lezotte, 1979; Edmonds, 1979; and Purkey and Smith, 1983.

With the Consortium for Chicago School Research joining the efforts of Designs for Change and the Chicago Panel on School Policy, foundations began to make more money available for research on improving schools. As school reform began to dominate the headlines of the daily newspapers, and foundations made funding available, researchers in the area's schools of education and related disciplinary departments began to undertake their own studies of Chicago's schools. In addition, scholars from across the country began to conduct their own inquiries or included Chicago in multicity studies of school reform efforts. A rich research literature on school reform, with Chicago at its center, began to develop.

In 1996, I left the Chicago Panel to become a research professor in the School of Education and Social Policy at Northwestern University. Continuing my focus on Chicago school reform, I was a member of the Steering Committee of the High School Redesign Project, convened by CEO Vallas. As the task force completed its work, members of Vallas's staff asked me to design a research effort to monitor the implementation of the secondary reforms. I presented the district with three options: an Extensive Evaluation including widespread use of ethnographers in most high schools, combined with widespread quantitative tracking of resource utilization and student achievement; a Minimal Monitoring model, relying heavily on statistical tracking of student achievement; and an Intermediate Investigation design incorporating some ethnography, some other forms of qualitative research, and extensive quantitative tracking of student performance. The district requested a full proposal based on the midlevel option, and the Center for Urban School Policy was established to conduct the three-year high school monitoring project.

Components of the New High School Design

The high school design adopted by the Reform Board of Trustees in March 1997 had two organizing themes: intensifying the academic "press" and enhancing personalism. The components of the design are organized under these two themes.

As shown in Table 1, six components of the design were focused on intensifying the pressure to improve academic performance of students, while four components were focused on enhancing personalism in the city's large and "anonymous" high schools.

Intensifying Academic Press

In 1996, the Reform Board of Trustees adopted new high school graduation requirements, boosting the number of required credits from 20 to 24 and adding additional years of English, science, mathematics, and social studies. The options for electives were reduced. Remedial courses did not bear credits toward graduation. These new requirements applied to students beginning high school in the fall of 1996. They were incorporated into the Design for High Schools adopted by the Reform Board the following spring.

Table 1

Components of the Chicago High School Design
Adopted March 1997

Intensifying Academic Press	Enhancing Personalism
1. Raised graduation requirements	1. Freshman Academics
2. Required enrollment in core courses	2. Advisories
3. Improving instruction via external partners	3. Small schools
4. Supporting principals via probation managers	4. Career academies
5. Core curriculum development and common semester exams	
6. High-stakes accountability: probation, reconstitution	

In addition to increasing graduation requirements, the Reform Board required that all high school students be scheduled for a course load that would produce six credits toward graduation every year, if the students passed all their courses. Students might be scheduled for additional periods of remedial support, if they required it. This was a major change for many inner-city high schools that had regularly scheduled many students, particularly freshmen, for no science courses and only remedial courses in reading and math.

As schools were placed on probation in 1996, they were provided with a list of about 40 approved potential external partners. Each potential partner had provided an agenda of assistance that they might provide schools; many partners focused on facilitation of improvement planning and implementation; some provided particular programs for schools; some focused more directly on improving classroom pedagogy and curriculum. In the first full year under probation, which was also the first year of implementation of the new high school design, schools were provided an average of about $100,000 to hire these external partners. In the second year, schools were expected to provide half the cost out of their discretionary funds (which averaged about $750,000 per probation high school). In the third year, if schools were still on probation, they would be required to fully fund the cost of their external partnership. This graduated increase in local cost of external partners was seen as an incentive to work quickly to improve so as to "get off" probation.

Probation managers were assigned to each school under that sanction, with the power to revise the school's improvement plan and to oversee school expenditures. Originally seen as part-time overseers of principals, the job description had already begun to morph toward a mentoring relationship by the time the high school redesign was being implemented. While proba-

tion managers still had to sign off on major expenditures, their role turned into being an advisor to the principal and a focus for staff planning and accountability for improvement.

Subject specific groups of teachers were convened in 1996 and 1997 to develop programs of study (common core curricula) for 11 key courses in English, math, science, and social studies. This was an effort to ensure that all high schools taught comparable content and skills in the basic courses in each discipline. At the same time, a second group of teachers was developing common end-of-semester exams for each of these courses: the Chicago Academic Standards Exams (CASE). While both groups of teachers were working from a common set of standards adopted by the district in 1996, there were some differences in emphasis and timing between the two sets of teacher-planners, causing some initial consternation among teachers across the city. After two years of pilot use, the CASE were incorporated into the marking system for students in the third year of implementation of the high school design.

Probation had already been established when the high school design was developed; the high-stakes accountability system was incorporated into the new design. During the summer of 1997, just as the new design was beginning to be implemented, seven high schools were reconstituted. In one school, 60 percent of the faculty were replaced; in another, as few as 20 percent were changed. Although no other high schools were reconstituted during the initial three years under the high school design, the district did develop a modified re-engineering design, modeled on the Toledo peer review plan developed in the late 1980s. In the fourth year of the design's implementation, five schools were placed into a new model called "intervention," which combined some aspects of reconstitution and of teacher pedagogical review.

Enhancing Personalism

Freshman Academies were designed to create smaller groups of students (about 100) in ninth grade who would share a group of core subject teachers. Ideally, these students and teachers would be colocated in adjoining rooms in a part of the building mostly devoted to first-year students. The idea had been piloted in one North Side high school during the year that the high school design was being formulated. The intent was for the students to get to know each other better than would be the case if they were roaming across the whole school to go from class to class, and for the teachers to have a common set of students to get to know, and to be able to discuss with each other the problems faced by individual students. It was intended that individual academies might undertake common interdisciplinary projects and engage in some social activities. After the first year of the high school design implementation, the central administration mandated that the academies continue into the sophomore year, renaming them Junior Academies.

Advisories were small groups of students — generally half a homeroom, or about 15 students — who would meet regularly (either weekly or daily) with a teacher or other adult on

the faculty of the school. The focus of advisories was to be on the social development of the students. The intent was to create a peer group within which difficult issues facing students could be discussed, and to create a trusting relationship between each student and at least one adult in the school. Advisories had been a successful feature of one of the better-known northern suburban high schools for decades.

The Chicago Public Schools had approved an initiative to create small schools within larger schools several years previously. The initiative had resulted from lobbying by a coalition of advocacy groups convened by a local public interest law firm. It was modeled after efforts begun in District 2 in New York City.[5] At least one external partner focused its attention on helping high schools create smaller schools within the host high school. Small schools were generally organized around a particular theme and were generally located in a specific part, or parts, of the host building. Small schools could be virtually autonomous, working collaboratively around facility-use issues, or they could be subunits still responsible to the school's administration.

Career academies were high schools that focused on preparing students for one of several different career arenas. Most career academies evolved out of former vocational high schools. Individual academies within career academy high schools could be largely autonomous (looking much like small schools with a particular career focus) or could be programs whose largest differences were in which vocational electives students chose.

What Kind of Research Is Appropriate to Measure Effectiveness of a Districtwide Reform Initiative?

Districtwide initiatives are quite different from initiatives that involve individual schools or groups of schools. To the extent that districtwide initiatives involve all schools in a district, they create opportunities not available in school-level initiatives, but they also create constraints on the research strategies that might be chosen to study them.

One immediate question researchers must confront is from whose perspective is effectiveness to be judged? An initiative might be judged very effective by the persons enacting the initiative, while others might find the initiative to be a diversion away from important concerns. In other cases, district administrators might find an initiative to be reaching its hoped-for results even while those charged with implementing the initiative are heavily criticizing it.

In the case of the Chicago high school redesign, there were at least four perspectives that were important to recognize. The school district had particular perspectives about what the

---

[5]See Fliegel, 1989; and Meier, 1995.

high school redesign was all about. But the district was composed of at least three subdivisions: the central administration; the leadership of local high schools; and the faculty, staff, and students in each school.

A second perspective was that of the larger research community in Chicago and across the nation. The research community shares many views on what good research is, but, by the late 1990s, a debate had begun within the research community about setting standards for rigorous work. The research community was also split ideologically. While some researchers focused on the pragmatic effects of the redesign initiative, approaching them with a relative neutrality about their intents, others came from a more critical, or neo-Marxist, perspective and saw the initiative as part of a large and sinister scheme for the accumulation and exercise of power.

A third perspective was the interest of the general public. By 1996, school reform had captured the attention of the news media for a decade or more. Members of the public who were not direct stakeholders in the public schools still were very interested in whether the school system was improving, and they saw the high school redesign effort as a key strategy for improvement.

Finally, the research contractor would have its own perspective. Sharing in each of the three previously mentioned perspectives, it would also have concerns about establishing goals that would be accomplished in conducting the research and about setting up the procedures for reaching those goals, including gaining the cooperation of the stakeholders who would be the subjects of the research. During the fall of 1997, a contract was negotiated between CPS and the Center for Urban School Policy at Northwestern University to conduct a three-year study of the implementation of the new high school design.

Potential Research Strategies

One of the constraints of studying districtwide initiatives which involve all or most schools in a school district is that there are no remaining schools to create an adequate control group for an experimental or quasi-experimental research design. Even schools which reject the initiative and do not implement its provisions are contaminated by their very rejection.

Some evaluation designs focus on the *process* of implementation of initiatives. Research designs built from this perspective devote much attention to how implementation proceeds, who becomes involved in the implementation, and what kind of involvement they exercise. Such designs intend to describe how well the initiative was implemented. Other evaluation designs are more concerned with the *outcomes* of the implementation of the initiative. Does the intended result occur? Does student achievement improve? In terms of the Chicago high school design, it was decided to create a research design which would produce findings about both the process of implementation and the effects on student achievement and behavior.

Many evaluation designs self-consciously adopt the perspective of the sponsoring agency seeking to implement the initiative being studied. The argument is that the researchers should suspend their own judgments and judge the initiative on the basis of whether or not it reached its own objectives. However, in the case of the Chicago high school design, as noted above, there were many different audiences and stakeholders, each of whom had an interest in the outcomes of the research being proposed. Therefore, it was decided to create a study design that would elicit many different perspectives on the initiative and to try to faithfully incorporate these different perspectives in the various reports produced by the research effort. Thus, the views of many different stakeholders were presented in the reports. Still, the researchers were compelled to judge all the various perspectives they uncovered and to make decisions about the weight of the evidence as gathered from diverse points of view.

A closely related issue is what measures will be utilized as the criteria for making judgments about the effectiveness of an initiative. In the case of the Chicago high school design, important components of the design used standardized reading and math scores to determine which aspects of the design would apply to which schools. Schools were placed on probation if less than 15 percent of enrolled students read or did math at or above the national norms (AANN). Schools were removed from probation if they raised the proportion AANN above 20 percent. The high schools chosen for reconstitution were the seven with the lowest proportion of students reaching the norms. Thus, the research design had to attend to these measures, the percentages reading and doing math at or above the national norms. But the very focus on these measures would have an effect on the implementation of the initiative, and the research had to also take into account the effect of the use of these measures on high schools seeking to implement the design. There were other goals, both quantitative and qualitative, that needed to be measured. And there were concerns brought by other researchers, the general public, and some stakeholders that were also worthy of investigation.

The CUSP Research Design

The leadership of the Chicago Public Schools asked the Center for Urban School Policy (CUSP) to develop a research proposal built on the Intermediate Investigations option. This request meant that not everything that might be done could be done. The scope of the research would be limited to about $600,000 per year for three years. Longer-term effects would be beyond the scope of the project. Similarly, not all schools could be given equal attention, if appropriate focus were paid to the most intrusive parts of the design. With a concern to produce findings about both the process of implementation and the effect of the initiative on student outcomes, a mixed-methods design was created that would focus on producing both quantitative and qualitative results.

Quantitative Measures

A number of different quantitative measures were adopted. Student achievement scores on standardized tests formed an important criterion for initiative implementation, as noted above, and so became an important component of the research design. CPS has utilized the Iowa Test of Basic Skills (ITBS) and its high school counterpart from Riverside Publishing, the Test of Achievement and Proficiency (TAP), for more than two decades. Scores on the TAP determined the probation status, and selection for reconstitution, for the city's high schools. Chicago schools are also subject to the state achievement tests. The state discontinued the test it had been using for a decade during the implementation of the high school design initiative; its replacement, the Illinois Standards Achievement Tests (ISAT), were initially available only in the elementary grades and, thus, played a smaller role as a measure of student achievement in this research effort.

Another form of achievement tests was developed during the implementation of the high school design, affecting classroom practice and student assessments as the years passed. Programs of study were developed during the first year of the design's implementation, and schools began to reorganize instruction around these curriculum guides starting in the second year of implementation. Pilot forms of the Chicago Academic Standards Exams began to be used in the spring of 1998, and school-level results were available to the research team and incorporated in its reports in 1999 and 2000. One big advantage of the CASE was that they were focused on the actual courses in which students were enrolled, rather than on more generalized assessments of reading and math ability. But these tests were imperfectly aligned with the programs of study and were somewhat inconsistent in difficulty between subjects and between years for the same subject. Rather than improve the quality of the CASE, CPS dropped their use in 2003, after only two years of public reporting.

Enrollment choice for high school attendance has a long history in Chicago. As early as 1991, more than half of Chicago high school students did not attend their neighborhood high school. Some attended magnet high schools; some attended vocational high schools that had no enrollment boundaries; others attended neighborhood high schools in other parts of the city. With the initiation of a system of sanctions on high schools with low levels of student achievement, and the subsequent public attention paid to school achievement levels, it became important to measure changes in school enrollments, as a measure of changes in enrollment choices exercised by the city's high school students.

It was also important to measure changes in course enrollment, grade point averages (GPA), and passing rates, given the new graduation requirements adopted by the Reform Board of Trustees and the mandate that all students be enrolled in a full schedule of credit-bearing courses.

Finally, it was important to understand patterns of behavior and opinions of the teachers being asked to implement the various components of the high school design. Therefore, each year, surveys were distributed to and collected from teachers in high schools experiencing the largest interventions under the high school design. Thus, teachers in all reconstituted high schools, those on probation, and those just above the probation cutoff were surveyed in each of the three years of the study.

Qualitative Measures

Two different qualitative approaches were incorporated into the design of the research. It was deemed imperative to understand in great detail what impact reconstitution had on the seven high schools subject to its provisions. Therefore, ethnographies of these seven schools were undertaken during the first two years of the study. The ethnographies were built on extensive interviewing among the leadership and general faculty and staff, observation of the classroom of every core subject teacher (English, math, social studies, science, world language), and intensive interviews with each teacher about the observed classroom.

Resources were not available to conduct ethnographies in all the high schools on probation, but it was important to understand how the high school design initiative was being implemented in each of those schools. Thus, in addition to surveys of the faculties of these schools, qualitative visitations were made to each of the 30 or so high schools on probation. These visitations entailed interviews with a wide range of school leaders and with a smaller selection of representative teachers. These interviews focused on the implementation of the high school design and on general operations of the school. In the second and third years, classroom observations were expanded to include probation high schools, focusing on a sample of teachers from the core subjects in the freshman year; all observed teachers were also interviewed about the observed class.

Graduated Focus on Schools with the Greatest Intervention

The constraint on resources implicit in the choice of the Intermediate Investigations option by CPS meant that the research design had to make strategic decisions about where to invest resources. The research design that was adopted focused more resources on schools experiencing the largest interventions, less resources on schools with moderate interventions, and very few resources on schools impacted mostly by changes in graduation requirements and mandated reorganizations of school structures. This meant that the research design focused its resources into the same schools in which the greatest reform resources of the school district were focused under the new high school design. More human resources were devoted to the seven high schools experiencing reconstitution, with a half-time ethnographer assigned to each of these schools, because it was felt to be critical to understand what happened in these schools as a re-

sult of the most intrusive component of the district's high-stakes accountability measures included in the high school design. Similarly, the next-largest investment of human resources was in creating a team of consultants to conduct qualitative visitations to high schools on probation. Each probation high school was visited once during the initial two years of the monitoring project. For high schools not on probation — but also not close to the national norms in student achievement — interviews were conducted with the principal; teachers were surveyed; and program, budget, and achievement data were analyzed. The third priority in resource allocation within the research design was in tracking quantitative data on student achievement, school and course enrollment, and use of resources. In all, about 20 persons participated on the research team at its highest staffing level.

During the third year of the monitoring project, the research design was changed, in consultation with CPS administrators, to shift resources from the reconstituted high schools to focus on an additional 12 schools undergoing the district's replacement strategy for reconstitution, called "reengineering." Under this revised design, these 12 schools plus the reconstituted high schools were visited three times during the year, with two sets of classroom observations conducted in each of these schools in the fall and the spring. The other high schools on probation continued to receive one visit and one set of classroom observations that year.

The strategy of the research plan was to build up to systemwide results from detailed reports of each school for each year. These reports followed a consistent format for all schools; the formats covered each of the major initiatives in the high school redesign plan. Each draft school report was sent to the major stakeholders in each school for their review and suggested amendment before being submitted as part of the annual report to the district. Because the research team was committed to including, as faithfully as possible, the multiple perspectives represented by the various constituencies in each school, concerns raised by the reviewers were always included in the final report on each school. These concerns might be balanced by contrary opinions or data raised by other stakeholders or by the judgments of the research team, but all were presented in the final report for each school.

The design of data collection and school visitations is presented in Table 2.

Assessing the Results of the New High School Design

When assessing the results of interventions in organizations, a simple four-cell cross-tab provides a conceptual frame. Interventions vary, generally in more complex ways than this frame implies, on two different scales: how well the intervention was implemented and what effect the intervention did have. The resulting four cells, in a simplified version of this framework, are shown in Table 3. Three of these boxes are easily understood and are the most likely result of new interventions. The best case is when the intervention is well implemented and the results hoped for

Table 2

Chicago High School Monitoring Project

Data Collection Plan

Nonprobation High Schools		Probation High Schools	
Schools Near National Norms (12)	Schools Near Probation, 15%-30% at Norms (15)	Schools on Probation, Not Reconstituted (32)	Reconstituted Schools (7)
	Program budget analysis Phone interviews	Plus school visits	Ethnographic case studies (classroom change) Plus school visits

NOTE: Quantitative data collection on staff, students, budget, program for all schools; analysis of data to be integrated into individual school reports on each school.

School Visitation Plan

Week 1	Week 2	Week 3	
Principal Interview	Key Actor Interviews	Teacher Interviews	
	Probation manager External partner Schools operations manager LSC chair Other administrator	In reconstituted schools (20) Department heads Small-school heads Affected faculty	In nonreconstituted schools
		Groups (2) Survey of all	Groups (2) Study of all

do occur; that is considered a successful intervention. The worst case is when the intervention is faithfully implemented, but the hoped-for outcomes do not result; that is an unsuccessful intervention. The intermediate case is when the intervention is not well implemented; advocates can then say the lack of result had to do with the faulty implementation, while critics can talk about why the intervention is difficult or impossible to implement. The fourth cell is the most anomalous. If the intervention is poorly implemented, it is not reasonable to think that the poor implementation "caused" the desired outcomes to occur. The researchers are left to try to explain what did lead to the desired outcomes, if it was not the intended intervention.

Table 3

Conceptual Frame for Assessing an Intervention

	Low Implementation	High Implementation
Low Outcomes	The intervention was not implemented well, and there were no significant changes in outcomes.	The intervention was faithfully implemented, but the hoped-for outcomes did not appear.
High Outcomes	The intervention was not implemented well, but the outcomes desired appeared anyway.	The intervention was implemented faithfully, and the hoped-for results occurred.

The new Chicago high school design fell into this fourth scenario. After three years of attempting to change the structure of high schools and the kind of instruction that occurred in their classrooms, the CUSP monitoring team found that very little change had actually occurred in classrooms or schools. At the end of the three-year research project, the high schools looked and felt very much as they had during the initial visits to these schools. Freshman/Junior Academies were mostly nonexistent or were present in name only. Advisories had been transformed from small groups focused on helping students overcome social development issues that might hinder their academic participation into tutoring periods focused on improving reading and math scores on standardized tests. The small schools and career academy programs that were established had little impact on the quality of instruction in classrooms.[6] A summary of major findings and recommendations is attached to this report as Appendix A.

The largest investments in the new high school design were focused on the schools that were also affected by the system's high-stakes accountability measures: schools that were on probation or reconstituted. It quickly became evident that reconstitution did not have the desired effect in the seven Chicago high schools where it was implemented. There was a fairly large turnover in faculty and staff in these high schools. Between 20 percent and 60 percent of the teachers were not rehired at their previous schools; five of the seven principals were not retained, and central-office administrators indicated that one of the retained principals would not have been, had there not been strong political interference in the decision.

---

[6]For a summary of the full report, see Hess and Cytrynbaum (2002). The report is also available on the CUSP Web site: See www.sesp.northwestern.edu/CUSP/ under "Research."

Thus, there was a significant amount of personnel change, though not the total change of reconstitution as practiced in San Francisco, the first major urban district to implement this strategy. However, interviews with principals, both at the beginning and at the end of the year, indicated that they did not think that they had significantly improved the quality of teaching in the school through the replacement process. Classroom observations of every core subject teacher in these seven schools indicated that the quality of instruction after reconstitution was quite low. (Because the monitoring project did not begin until after staffs had been reconstituted, the research team had no framework for comparing the current performance of teachers to prior performance.) Teachers' perceptions were that the strategy had been a sham and had simply resulted in "turning over the applecart," with a needless interruption of their colleagues' careers. Principals indicated that some of the problem was rooted in bad personnel decisions that they had made under a very constricted process, which was limited to interviewing potential rehires or new hires without the opportunity of viewing the candidates actually teaching. But of more importance, principals said, was the very limited number of high-quality teachers in the pool of potential hires. Indeed, it became evident that, at the end of the process, reconstituted high schools were hiring their colleagues' rejects just to be able to fill all classroom positions before school started. Thus, reconstitution did not produce dramatically improved faculty quality, and it undermined whatever school morale had existed in these schools prior to reconstitution. It became clear that any "teacher replacement" strategy, such as reconstitution, can only be successful if there is a large enough supply of high-quality replacement candidates. As a result of the monitoring team's interim report on reconstitution — confirmed by central-office administrators working with these schools — and in the face of grievances filed by the teachers union, CPS abandoned the reconstitution process and began designing an alternative teacher assessment and improvement process called "reengineering." However, this plan, which was built on the foundation of the Toledo Peer Counseling program, was never fully implemented, and, after two years, it was succeeded by yet another effort, called "intervention."[7]

The next-largest investment of resources under the high school design was for external partners to work with faculties of schools on probation. Nine external partners were selected by the 39 high schools on probation, from a list of approved providers made available from the central office. Four of these partners served two-thirds of the schools, while the other five served only a few schools each. Only 13 of the schools kept the same partner for the studied three-year implementation period. Most schools changed partners at least once, and several changed partners twice or three times. Teacher resistance to the assistance of the external partners was high in many schools, particularly in schools where there had not been a prior relationship with the partner. Observations in more than 800 classrooms by the research team indicated

---

[7]For a fuller account of the effects of reconstitution, see Hess, 2003; the report can also be found on the CUSP Web site: See www.sesp.northwestern.edu/CUSP/ under "Research."

that instruction was generally at a very shallow level and that, in half the classrooms, five or fewer students could be said to be fully engaged in the lesson. On the basis of classroom observations and interviews with teachers after every observation, the research team concluded that there were three major obstacles to improving instruction:

- Some teachers did not know their subject matter.

- Some teachers who knew their subject did not know how to get it across.

- Many of the weakest teachers did not believe that their students could learn their subject material.

The external partners tended to focus on the second of these obstacles and generally were able to get compliance of teachers in adopting peripheral pedagogical strategies, such as the use of "bell-ringers" to start classes, a focus on vocabulary enhancement by schoolwide use of a "word of the day," and the utilization of graphic organizers to help students visualize an array of data or concepts. But their efforts had little impact on the depth of instruction or extending the number of students engaged in classroom lessons.[8] Correspondingly, when improvement of student achievement was measured between students' ninth-grade reading scores and their scores in tenth or eleventh grade, most external partners averaged about one more student reading at the national norms than had done so in ninth grade for each school for each year. This was very minimal improvement in the primary measure of student achievement in use in the district. The CUSP research team felt that this very limited impact on student achievement reflected teacher resistance to the external partners and the partners' neglect of the third obstacle, teachers' beliefs about their students' abilities.

There was little evidence available to the research team that any significant changes had occurred in Chicago's high schools except on two of the ten primary components of the new high school design. The raised high school graduation requirements — together with the mandate that all students be programmed for a full academic, credit-bearing load — did result in students taking more academic courses, particularly more math and science courses. Thus, on most aspects of the new high school design, implementation should fairly be assessed as very low. Given low implementation of the intervention, most observers would expect (especially given the intransigence of student achievement levels nationally) that there would be little improvement in student achievement in Chicago's high schools.

However, that was not the case. Student achievement went up on many measures between 1996, the year before the policy began to be implemented, and 2000, the final year of the

---

[8]CUSP chose to use the metaphor of "depth of instruction"; others have chosen the metaphor of "higher-order thinking skills." The concepts are similar.

study. Student achievement on standardized reading and math tests — the district's primary indicator of student achievement — improved both across the system and in the schools on probation. As noted above, students were taking more core subject courses and more advanced math and science courses, and their average GPAs were improving. Student attendance was improving marginally. Although these improvements in student outcomes did not happen in every school and many schools saw improvement in one year followed by deterioration in the next, most schools were seeing their reported achievement scores rising. However, as the research team reported to CEO Vallas at one point: "The good news is that student achievement in high schools is up. The bad news is that the high schools are not to blame for it!"

Indeed, principals, administrators, external partners, and lead teachers were quick to claim credit for student achievement improvement, even when the research team could find no evidence of the improvements that they claimed were responsible for student learning gains. This left the research team in a quandary: What did account for higher student achievement, if it was not the result of changes in school structure or classroom instruction as envisioned in the new high school design?

A careful analysis of the prior achievement records of the city's 100,000 annually enrolled high school students showed a pattern that did make sense of the improved achievement despite no significant changes in the city's high schools. By tracking the prior achievement of cohorts of students, the team was able to demonstrate that successive cohorts of entering freshmen left eighth grade performing at higher and higher levels. This was in part because the average achievement scores of eighth-graders were improving, and in part it reflected the effects of the system's elementary school promotion gates policy, which kept some of the poorest-performing elementary school students from entering high school at all or kept them out of the pool of students whose test scores counted under the system's accountability system. When these higher-performing successive cohorts of students were tested in ninth grade and again in either tenth or eleventh grade, they naturally scored at higher levels of these tests than had their predecessors. To the extent that the standardized tests also reflected that entering ninth-graders were better prepared for high school, better GPAs in more difficult courses were also understandable. In short, higher high school achievement was apparently resulting from improved achievement in the city's elementary schools, not the result of improvements in the city's 20 high schools.

The Continuing Challenge

Improving urban high schools continues to be a formidable challenge in the United States. While there are isolated success stories, the history of restructuring existing urban high schools continues to be primarily a story of failure. Bright spots are more frequently found among newly created schools, but much of that evidence is idiosyncratic. There is little evi-

dence yet of reforms going to scale and encompassing most of the high school students in urban school districts.

As the CUSP research team considered the next steps toward high school reform in Chicago, it focused on three efforts. The first was to encourage a stronger effort to actually implement the structural reorganizing that made up the enhancing personalism part of the design: junior academies, advisories, small schools, and career academies. The team recognized that there were low levels of teacher "buy-in" relative to these restructuring efforts that had to be better understood and seriously considered before renewed efforts would be likely to be successful. But the team felt that these aspects of the design still merited stronger implementation efforts.

The second challenge involved the effort to improve classroom instruction. It was evident from three years of classroom observation that there was lots of room for instructional improvement in the high schools of Chicago. But if professional development efforts were limited to improving teachers' knowledge of their subject matter and improving their pedagogical skills and repertoire, the prospects for success were likely to be far below what was necessary. This is not to say that efforts directed at content and pedagogical knowledge and skill improvement are not important and valuable; the team believed that they were. But such efforts would not be sufficient, in Chicago. The deeper issue that must be addressed is the beliefs of teachers about the ability of their students. During the three years under study, higher proportions of teachers each year (growing from 63 percent to 70 percent) said that their students were unprepared to tackle the subjects the teacher was teaching! This declining confidence of teachers about their students' ability occurred at the same time that the entering achievement levels of freshmen were constantly improving.

The research team came to understand that many Chicago teachers had 15 or 20 years' experience of students not performing well academically. Teachers had developed mechanisms to cope with that career-long disappointment. Some teachers focused on the students who did seem prepared and who were eager to do assignments and to participate in class. These teachers were rewarded with individual success stories of students who performed well in high school and went on to further success in higher education or employment, while the majority of their students continued to flounder. Other teachers simply blamed the elementary schools, or the families of their students, or the persistent poverty that characterized the communities from which their students come. These teachers' beliefs about their students were rooted in years of experience; these beliefs would not be changed by admonitions to teachers to "raise their expectations" or by diatribes about the racism or classism that their beliefs reflected.

But the research team was also mindful that a third of the teachers who were observed did teach students to think deeply, and a quarter were capable of engaging most of the students in their classes. There are good urban high school teachers in the city's probation high schools.

But, as presently located, these teachers do not constitute a critical mass that can change the faculty culture of high schools to expect that all teachers can have similar success with their students. The result is that there are no powerful models of successful urban high schools in Chicago.[9] It might behoove the city to undertake to aggregate enough of these successful urban teachers in one high school to develop such a model in a typical urban neighborhood with a typical inner-city enrollment of students.

The third concern of the research team was with the narrowing of improvement efforts that had resulted from using standardized reading and math tests as the primary measure of a school's student achievement. By focusing attention for accountability primarily on reading, external partners shifted their focus to follow suit, frequently while bemoaning the abandonment of broader efforts through which they had qualified to be on the approved list. The focus on reading also meant teachers of science, social studies, languages, and other subjects felt that they were not responsible for their school's achievement levels and were largely exempt from the need to improve the level of instruction. English and math teachers in schools with rising achievement scores on standardized tests felt that these results vindicated their current practices, and thus they became less open to efforts to improve pedagogy. Meanwhile, students' scores on the CASE — the district's subject specific semester exams — were abysmal in all subjects except English I and II. While there were serious problems with the consistency and rigor of the CASE, the CUSP team recommended moving away from standardized reading and math tests and toward greater utilization of the subject-specific exams in the district's accountability program, in order to broaden the effort to improve pedagogy in the city's lowest-achieving schools. The team also recommended a thorough review and revision of the CASE by testing experts.

Other Research on Chicago's High Schools

While the CUSP monitoring study was the largest research effort focused on Chicago's high schools, there were other significant projects also under taken during the late 1990s. In March 2001, the Consortium on Chicago School Research sponsored a forum on Research in Chicago's High Schools, which became the occasion for the public release of CUSP's final report. The papers presented at that conference were collected into a volume published in 2002.[10] While many of these studies were small-scale qualitative studies, consortium staffers Shazia

---

[9]At the elementary school level, Chicago has developed a number of schools that serve exclusively low-income and minority students whose achievement matches that of students across the country. These schools stand as existence proof in the city that low-income students can learn at levels commensurate with their peers nationally.

[10]Lee, 2002.

Miller and Elaine Allensworth presented data on high school student performance between 1993 and 2000 that complemented the CUSP study.[11]

Miller and Allensworth's highly statistical study confirmed the rising scores of entering freshmen, which had also been reported in the CUSP study. Their data show a small rise in the graduation rate, from 51.0 percent to 53.6 percent of students who had been in the system as 13-year-olds (end of seventh grade for most students), and a corresponding decline in dropout rates, most of which was attributed to the improved achievement levels of entering students. They reported that more students were remaining on-track for graduation after four years than had been previously been the case and that more students were passing the college-track courses. They confirmed data from the CUSP study that more students were attempting and passing credit-bearing courses. Specifically, more students were taking and passing the Algebra I–Geometry sequence. They further contended that student reading and math scores on the TAP were rising even more than what might be accounted for by the improvement of prior achievement by entering freshmen. They noted the sharp decline in students reading in the bottom quartile among students entering high school after the establishment of the eighth-grade promotion gate for the 1996-97 school year.

A bibliography of research on Chicago schooling is attached to this report as Appendix B.

References

Brookover, Wilbur B., and Larry W. Lezotte. 1979. *Changes in School Characteristics Coincident with Changes in Student Achievement.* East Lansing: Michigan State University.

Bryk, Anthony S., John Q. Easton, David Kerbow, Sharon G. Rollow, and Penny A. Sebring. 1993. *A View from the Elementary Schools: The State of Reform in Chicago.* Chicago: Consortium on Chicago School Research.

Chicago Public Schools. 1997. *Design for High Schools.* Chicago: Chicago Public Schools.

Edmonds, Ronald R. 1979. "Effective Schools for the Urban Poor." *Educational Leadership* 37 (October): 15-18.

Fliegel, Sy. 1989. "Parental Choice in East Harlem Schools." Pages 95-112 in Joe Nathan (Ed.), *Public Schools by Choice.* Minneapolis: Institute for Learning and Teaching.

Hess, G. Alfred, Jr. 1991. *School Restructuring, Chicago Style.* Thousand Oaks, CA: Corwin Press.

Hess, G. Alfred, Jr. 2003. "Reconstitution — Three Years Later: Monitoring the Effect of Sanctions on Chicago High Schools." *Education and Urban Society* 35, 3: 300-327. Also available at www.sesp.northwestern.edu/CUSP/ under "Research."

---

[11]Miller and Allensworth, 2002.

Hess, G. Alfred, Jr., and Solomon Cytrynbaum. 2002. "The Effort to Redesign Chicago High Schools: Effects on Schools and Achievement." Pages 19-49 in Valerie Lee (Ed.), *Research on High School Reform Efforts in Chicago.* Chicago: Consortium on Chicago School Research. Also available at www.sesp.northwestern.edu/CUSP/ under "Research."

Lee, Valerie E. (Ed.). 2002. *Reforming Chicago's High Schools: Research Perspectives on School and System Level Change.* Chicago: Consortium on Chicago School Research.

Meier, Deborah. 1995. *The Power of Their Ideas: Lessons for America from a Small School in Harlem.* Boston: Beacon Press.

Miller, Shazia Rafiullah, and Elaine M. Allensworth. 2002. "Progress and Problems: Student Performance in CPS High Schools, 1993 to 2000." Pages 51-87 in Valerie E. Lee (Ed.), *Reforming Chicago's High Schools: Research Perspectives on School and System Level Change.* Chicago: Consortium on Chicago School Research.

Purkey, Stan C., and Marshall S. Smith. 1983. "Effective Schools: A Review." *Elementary School Journal* 81, 1 (1983): 426-452.

Appendix A

The Effort to Redesign Chicago High Schools:
Effects on Schools and Achievement

G. Alfred Hess, Jr., and Solomon Cytrynbaum
Center for Urban School Policy
School of Education and Social Policy, Northwestern University

Summary of Major Findings

Overarching Finding

High school students are scoring higher on standardized tests than were their predecessors, but that is not because schools have dramatically changed what they are doing with their students.

The good news is that academic achievement in the city's high schools has improved significantly since the baseline year of 1996. The proportion of students reading at or above the national norms has risen from 20.5 percent in 1996 to 35.3 percent in 2000. Correspondingly, the proportion doing math at the norm has risen from 21.7 percent in 1996 to 45.0 percent in 2000. Thus, more than a third of Chicago's high school students read at the national norm, and almost half do math at that level. However, the data included in this report indicate that, despite the hard work of many, little significant change has happened in the city's lower-performing high schools. Advisories were held weekly in most probationary high schools, but their content was diluted. Junior Academies were fully implemented in only seven of these high schools. Small schools were in place in four of them, and career academies were in 12. Virtually all the improvement in test scores reflects changes in preparedness of students entering the city's high schools. There is little reason to think that students' social development has been significantly improved since 1997 either.

Lessons from Implementing the Design for High Schools

1. **The context of high school education has been changed. The focus has shifted from warehousing and managing the behavior of kids to a focus on student learning.**

   - There is a new focus on the academic performance of high schools and their students.

   - Accountability has helped to give weight to efforts in individual high schools on probation to instill higher curriculum standards.

   - Teacher interdependence has been fostered, and teachers' practice has become more public.

2. **The focus on improving Chicago's lowest-performing high schools, along with a concentration of resources to support those schools, has been a grand effort to enhance social mobility.**

- Schools enrolling predominantly low-income and minority students have been held to the same standards as those with more advantaged students.

- Additional resources were focused into the schools serving the least advantaged students.

3. **Students have been asked to work harder and to invest much more in their own education. School staffs have also been asked to work harder.**

- Students are required to take credit-bearing courses, and they face expanded testing of their achievement (five-week exams, the CASE, and standardized tests).

- More significant, however, than the efforts in high schools to ask students to work harder has been the prior efforts in elementary schools to require better performance to gain entrance to high school.

- Teachers and principals have additional responsibilities, coordinate more, and work longer hours.

4. **Outsiders have been engaged to assist in the effort to improve Chicago's lowest-performing high schools.**

- The use of probation managers and external partners to assist schools reverses the traditional posture of excluding outside influences.

- Schools served by external partners have added to the number of students reading at the norms between ninth grade and subsequent tests.

- Teachers appreciated the help of external partners but thought that they contributed little to improvement in test scores; they thought that teachers had contributed most.

5. **The Design for High Schools envisioned a significant effort to change teacher behaviors, both pedagogically and in their relationships with their students. However, as large as the effort was (with costs of approximately $100,000 per high school per year), it did not prove to be intensive enough.**

- Most teachers (58.3 percent) taught very shallowly in 2000; 48 percent narrowed instruction to five or fewer students.

- Appropriate levels of content were taught in two-thirds (68 percent) of observed classrooms. CUSP staff observed three major problems in the system's weaker high school teachers.

- Some of these teachers did not know their subject matter very well.

- Some of these teachers did not know how to get their subject matter across to their students.

- A third group of teachers, which might include some from the first two groups, did not believe that their students were capable of learning the material and skills included in the Chicago Academic Standards.

- Despite rising test scores of entering freshmen, teachers claimed that student preparation was deteriorating. More teachers reported low morale in their school than in 1998.

- Regardless of the appropriateness of teachers' responses, a demoralized teacher workforce is a major challenge to further efforts to improve Chicago's high schools.

6. **The realignment of Special Education instruction, following the settlement of the Corey H. litigation, has resulted in the intersection of two different reforms that has made both more difficult to achieve.**

- The number of Special Education students in Chicago high schools has increased by 2,381 — a 20.7 percent increase between 1996 and 2000.

- There were many more Special Education students mainstreamed in regular education classrooms as a result of the Corey H. settlement.

- Special Education students were disproportionately enrolled in the city's lowest-performing schools.

- Many teachers confronted classrooms in which between a quarter and a third of their students were Special Education students; the teachers were not up to the challenge of both improving their quality of instruction and learning how to differentiate instruction to meet a wide range of student preparation in the same classroom.

7. **The focus on reading — while necessary to enforce accountability and to change the context of valuing student learning — has proved ineffective and is diverting attention from the substance of the core curriculum.**

- The focus on reading has had only minimal effect on increasing the number of better-performing readers, once students enter high school. These efforts account for only a minuscule part of the improvement in achievement scores.

- The focus on reading has diverted time and energy away from the prescribed curricula of core courses at exactly the same time that programs of study and the Chicago Academic Standards Exams have been established with very broad content coverage requirements.

- An unanticipated effect of rising reading and math scores in high schools has been an undermining of teachers' perceptions of the need for change.

- In the absence of substantial evidence of the effectiveness of the reading strategy and with the development of more objective measures of subject-specific learning, it is now time to shift the focus of accountability onto the CASE and away from the TAP and the forthcoming Prairie State exams.

  - A major component in teacher dissatisfaction with the CASE and with the underlying programs of study relates to the tremendous breadth of material potentially encompassed by each of these exams.

  - This "breadth over depth" focus of the CASE reinforces the shallowness of instruction that has been highlighted as characterizing the majority of classrooms described in this report.

Recommendations

1.   Focus more attention on restructuring to enhance personalism.

A rededication to building advisories and Junior Academies, and to supporting small schools and career clusters within career academies, is needed.

2.   Intensify efforts at teacher development.

The current efforts are not intense enough to change teachers' beliefs. Potential models include the Schenley Training Center approach pioneered in Pittsburgh in the 1980s and the Lead Teacher model being implemented in Manley High School.

3.   Shift the focus of accountability back to assessing learning in the core subjects.

Shift away from using standardized tests of reading to the CASE in core subjects. This will require attention to the year-to-year reliability of the CASE and to shifting their emphasis from breadth of coverage to depth of understanding, with a comparable shift in the programs of study. Additional assessment arenas might be graduation rates and passing rates, as long as the CASE provide common, objective measures among schools.

Appendix B

Bibliography on Chicago School Reform

NOTE: An asterisk (*) denotes a high school focus.

Books

Bryk, Anthony S., and Barbara Schneider. 2002. *Trust in Schools: A Core Resource for Improvement.* New York: Russell Sage Foundation.

Bryk, Anthony S., Penny Bender Sebring, David Kerbow, Sharon Rollow, and John Q. Easton. 1998. *Charting Chicago School Reform: Democratic Localism as a Lever for Change.* Boulder, CO: Westview Press.

Hess, G. Alfred, Jr. 1995. *Restructuring Urban Schools: A Chicago Perspective.* New York: Teachers College Press.

Hess, G. Alfred, Jr. (Ed.). 1994. *Outcomes of Chicago School Reform*, a special issue of *Education and Urban Society* 26, 3 (May).

Hess, G. Alfred, Jr. 1991. *School Restructuring, Chicago Style.* Thousand Oaks, CA: Corwin Press.

Kyle, Charles L., and Edward R. Kantowicz. 1992. *Kids First — Primero Los Ninos: Chicago School Reform in the 1980s.* Springfield, IL: Illinois Issues.

Lee, Valerie E. (Ed.). 2002. *Reforming Chicago's High Schools: Research Perspectives on School and System Level Change.* Chicago: Consortium on Chicago School Research.

Pre-Reform Studies

Chicago Panel on Public School Policy and Finance. 1988. *Illegal Use of State Chapter I Funds*, Chicago: Chicago Panel on Public School Policy and Finance.

Designs for Change. 1985. *The Bottom Line: Chicago's Failing Schools and How to Save Them.* Chicago: Designs for Change.

Hess, G. Alfred, Jr. 1984. "Renegotiating a Multicultural Society: Participation in Desegregation Planning in Chicago." *Journal of Negro Education* 53: 132-146.

Hess, G. Alfred, Jr., Denise O. Green, Anne E. Stapleton, and Olga Reyes. 1988. *Invisibly Pregnant: Teenage Mothers and the Chicago Public Schools.* Chicago: Chicago Panel on Public School Finances.

Hess, G. Alfred, Jr., and James Greer. 1987. *Bending the Twig: The Elementary Years and Dropout Rates in the Chicago Public Schools.* Chicago: Chicago Panel on Public School Finances.

*Hess, G. Alfred, Jr., and Diana Lauber. 1985. *Dropouts from the Chicago Public Schools.* Chicago: Chicago Panel on Public School Finances.

Hess, G. Alfred, Jr., Arthur Lyons, and Louis Corsino. 1989. *Against the Odds: The Early Identification of Dropouts.* Chicago: Chicago Panel on Public School Policy and Finances.

Hess, G. Alfred, Jr., and Hannah Meara. 1984. *Teacher Transfers and Classroom Disruption.* Chicago: Chicago Panel on Public School Finances.

Hess, G. Alfred, Jr., and Christina Warden. 1988. "Who Benefits from Desegregation Now?" *Journal of Negro Education* 57: 536-551.

*Hess, G. Alfred, Jr., Emily Wells, Carol Prindle, Beatrice Kaplan, and Paul Liffman. 1986. *"Where's Room 185?" How Schools Can Reduce Their Dropout Problem.* Chicago: Chicago Panel on Public School Finances.

*Kyle, Charles. 1984. *Aspira Chicago Hispanic Dropout Study.* Chicago: Aspira of Illinois.

Moore, Donald, and Sheila Radford Hill. 1982. *Caught in the Web: Misplaced Children in Chicago's Classes for the Mentally Retarded.* Chicago: Designs for Change.

Reform, Phase I Studies (1990 - 1995)

See, under "Books," Hess (Ed.), 1994; Hess, 1995; Bryk Sebring, Kerbow, Rollow, and Easton, 1998.

Bennett, Albert L., Anthony S. Bryk, John Q. Easton, David Kerbow, Stuart Luppescu, and Penny A. Sebring. 1992. *Charting Reform: The Principal's Perspective.* Chicago: Consortium on Chicago School Research.

Bryk, Anthony S., Paul E. Deabster, John Q. Easton, Stuart Lupescu, and Yeow M. Thum. 1994. "Measuring Achievement Gains in the Chicago Public Schools." *Education and Urban Society* 26 (3): 306-319.

Bryk, Anthony S., John Q. Easton, David Kerbow, Sharon G. Rollow, and Penny A. Sebring. 1993. *A View from the Elementary Schools: The State of Reform in Chicago.* Chicago: Consortium on Chicago School Research.

Designs for Change. 1993. *Chicago Achievement Scores and Graduation Rates Rise Since Reform.* Chicago: Designs for Change.

Easton, John Q. 1989. *Teacher Attitudes Toward School Reform.* Chicago: Chicago Panel on Public School Policy and Finance.

Easton, John Q., Anthony S. Bryk, Mary E. Driscoll, John G. Kotsakis, Penny A. Sebring, and Aurie J. van der Ploeg. 1991. *Charting Reform: The Teachers' Turn.* Chicago: Consortium on Chicago School Research.

Easton, John Q., Susan L. Flinspach, Darrel J. Ford, Jesse Qualls, Susan Ryan, and Sandra L. Storey. 1991. *Decision Making and School Improvement: LSCs in the First Two Years of Reform.* Chicago: Chicago Panel on Public School Policy and Finance.

Flinspach, Susan L., and Susan P. Ryan. 1994. "Diversity of Outcomes: Local Schools Under School Reform." *Education and Urban Society* 26 (3): 292-305.

Flinspach, Susan L., and Susan P. Ryan. 1991. *Vision and Accountability in School Improvement Planning.* Chicago: Chicago Panel on Public School Policy and Finance.

Ford, Darrel J. 1991. *The School Principal and Chicago School Reform: Principals' Early Perceptions.* Chicago: Chicago Panel on Public School Policy and Finance.

Hess, G. Alfred, Jr. 1994. "School Based Management as a Vehicle for School Reform." *Education and Urban Society* 26, 3 (May): 203-219.

Hess, G. Alfred, Jr. 1994. "The Changing Role of Teachers: Moving from Interested Spectators to Engaged Planners." *Education and Urban Society* 26, 3 (May): 248-263.

Hess, G. Alfred, Jr. 1994. "Chicago School Reform: A Response to Unmet Needs of Students at Risk." Pages 207-228 in Robert J. Rossi (Ed.), *Schools and Students at Risk.* New York: Teachers College Press.

Hess, G. Alfred, Jr. 1994. "New Roles Under School-Based Management: The Chicago School Reform Act." Pages 55-72 in Robert J. Yinger and Kathryn Borman (Eds.), *Restructuring Education: Issues and Strategies for Communities, Schools, and Universities.* Cresskill, NJ: Hampton Press.

Hess, G. Alfred, Jr. 1992. "Midway Through School Reform in Chicago." *International Journal of Education* 1 (13): 270-284.

Hess, G. Alfred, Jr.. and John Q. Easton. 1992. "Who's Making What Decisions: Monitoring Authority Shifts in Chicago School Reform." Pages 157-176 in G. Alfred Hess, Jr. (Ed.), *Empowering Teachers and Parents: School Restructuring through the Eyes of Anthropologists.* Westport, CT: Bergin and Garvey.

Hess, G. Alfred, Jr., Susan L. Flinspach, and Susan P. Ryan. 1993. "Case Studies of Chicago Schools Under Reform." *New Directions for Program Evaluation* 59 (Fall): 43-55.

Hess, G. Alfred, Jr., and John Q. Easton. 1994. "Monitoring the Implementation of Radical Reform: Restructuring the Chicago Public Schools." Pages 221-248 in Kathryn M. Borman and Nancy P. Greenman (Eds.), *Changing American Education: Recapturing the Past or Inventing the Future?* Albany, NY: SUNY Press.

Sebring, Penny A., Anthony S. Bryk, John Q., Easton Stuart Luppescu, Yeow M. Thum, Winnifred Lopez, and BetsAnn Smith. 1995. *Charting Reform: Chicago Teachers Take Stock.* Chicago: Consortium on Chicago School Research.

Reform, Phase II Studies (1996 - 2003)

See, under "Books," Lee, 2002.

*Allensworth, Elaine M., and Shazia R. Miller. 2002. *Declining High School Enrollment: An Exploration of Causes.* Chicago: Consortium on Chicago School Research.

Bryk, Anthony S., Jenny K Nagaoka, and Fred M. Newmann. 2000. *Chicago Classroom Demands for Authentic Intellectual Work.* Chicago: Consortium on Chicago School Research.

Bryk, Anthony S., Yeow M. Thum, John Q. Easton, and Stuart Luppescu. 1998. *Academic Productivity of Chicago Public Elementary Schools.* Chicago: Consortium on Chicago School Research.

Designs for Change. 1998. *What Makes These Schools Stand Out? Chicago Elementary Schools with a Seven-Year Trend of Improved Reading Achievement.* Chicago: Designs for Change.

Finnigan, Kara, and Jennifer O'Day. 2003. *External Support to Schools on Probation: Getting a Leg Up?* Chicago: Consortium on Chicago School Research.

Hanson, Matthew R., and Donald R. Moore. 2003. *Rachel Carson Elementary School: An Exemplary Urban School That Teaches Children to Read.* Chicago: Designs for Change.

Hart, Holly, Elaine M. Allensworth, Doug Lauen, and Robert M. Gladden. 2002. *Educational Technology: Its Availability and Use in Chicago's Public Schools.* Chicago: Consortium on Chicago School Research.

*Hess, G. Alfred, Jr. 2003. "Reconstitution — Three Years Later: Monitoring the Effect of Sanctions on Chicago High Schools." *Education and Urban Society* 35, 3: 300-327.

*Hess, G. Alfred, Jr. 2002. *Teaching and Learning Under Chicago School Reform: Lessons from High School Math and Science Classrooms.* Evanston, IL: Center for Urban School Policy.

*Hess, G. Alfred, Jr., and Solomon Cytrynbaum. 2001. *The Effort to Redesign Chicago's High Schools: Effects on Schools and Achievement.* Evanston, IL: Center for Urban School Policy.

Hess, G. Alfred, Jr. 2001. "Accountability in Chicago." Pages 339-374 in Diane Ravitch (Ed.), *Brookings Papers on Education Policy 2001.* Washington: Brookings Institution Press.

Hess, G. Alfred, Jr. 2000. *Examining Achievement Differences in Predominantly African American Schools.* Evanston, IL: Center for Urban School Policy.

Hess, G. Alfred, Jr. 1999. "Expectations, Opportunity, Capacity, and Will: The Four Essential Components of Chicago School Reform." *Educational Policy* 13, 4 (September): 494-517.

Hess, G. Alfred, Jr. 1999. "Understanding Achievement (and Other) Changes Under Chicago School Reform." *Educational Evaluation and Policy Analysis* 21, 1 (Spring): 67-83.

Lee, Valerie E., Julia B. Smith, Tamara E. Perry, and Mark A. Smylie. 1999. *Social Support Academic Press, and Student Achievement: A View from the Middle Grades in Chicago.* Chicago: Consortium on Chicago School Research.

*Miller, Shazia R., Elaine M. Allensworth, and Julie R. Kochanek. 2002. *Student Performance: Course Taking, Test Scores, and Outcomes.* Chicago: Consortium on Chicago School Research.

Miller, Shazia R., and Robert M. Gladden. 2002. *Changing Special Education Enrollments: Causes and Distribution Among Schools.* Chicago: Consortium on Chicago School Research.

Moore, Donald R. 2003. *An Alarming Percentage of Hispanic Youth in the Chicago Metro Area Are Dropouts and Jobless.* Chicago: Designs for Change.

Moore, Donald R. 2001. "Changing the Ground Rules." *Shelterforce* (July/August).

Moore, Donald R. 2000. *New Data About Chicago's Grade Retention Program Provides Further Proof That Neither Retention Nor Social Promotion Works — Research-Based Alternatives to Both Retention and Social Promotion Should Be Carried Out.* Chicago: Designs for Change.

Moore, Donald R., and Matthew Hanson. 2001. *National Test Experts Find Basic Flaws in Chicago's Use of Iowa Tests to Make Critical Decisions About Students and Schools.* Chicago: Designs for Change.

Moore, Donald R., and Matthew Hanson. 2001. *Chicago Test Score Research Shows That School-Level Initiative Brings the Largest Sustained Reading Gains: Evidence Indicates the Need for "Phase Three" of Chicago School Reform.* Chicago: Designs for Change.

Moore, Donald R., and Gail Merritt. 2002. *Chicago's Local School Councils: What the Research Says.* Chicago: Designs for Change.

Newmann, Fred M., Anthony S. Bryk, and Jenny K. Nagaoka. 2001. *Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence.* Chicago: Consortium on Chicago School Research.

Newman, Fred M., Gudelia Lopez, and Anthony S. Bryk. 1998. *The Quality of Intellectual Work in Chicago Schools: A Baseline Report.* Chicago: Consortium on Chicago School Research.

Newmann, Fred M., BetsAnn Smith, Elaine Allensworth, and Anthony S. Bryk. 2001. *School Instructional Program Coherence: Benefits and Challenges.* Chicago: Consortium on Chicago School Research.

Roderick, Melissa, Anthony S. Bryk, Brian A. Jacob, John Q. Easton, and Elaine Allensworth. 1999. *Ending Social Promotion: Results from the First Two Years.* Chicago: Consortium on Chicago School Research.

Roderick, Melissa, Mimi Engel, and Jenny Nagaoka. 2003. *Ending Social Promotion: Results from Summer Bridge.* Chicago: Consortium on Chicago School Research.

Roderick, Melissa, Jenny K. Nagaoka, Jen Bacon, and John Q. Easton. 2000. *Update: Ending Social Promotion.* Chicago: Consortium on Chicago School Research.

Ryan, Susan P., Anthony S. Bryk, Gudelia Lopez, Kimberly P. Williams, Kathleen Hall, and Stuart Luppescu. 1997. *Charting Reform: LSCs-Local Leadership at Work.* Chicago: Consortium on Chicago School Research.

Sebring, Penny B., Anthony S. Bryk, Melissa Rocerick, Erik Camburn, Stuart Luppescu, Yeow M. Thum, BetsAnn Smith, and Joseph Kahne. 1996. *Charting Reform in Chicago: The Students Speak.* Chicago: Consortium on Chicago School Research.

Shipps, Dorothy, Karin Sconzert, and Holly Swyers. 1999. *The Chicago Annenberg Challenge: The First Three Years.* Chicago: Consortium on Chicago School Research.

Smith, BetsAnn. 1998. *It's About Time: Opportunities to Learn in Chicago's Elementary Schools.* Chicago: Consortium on Chicago School Research.

Smith, Julia B., BetsAnn Smith, and Anthony S. Bryk. 1998. *Setting the Pace: Opportunities to Learn in Chicago's Elementary Schools.* Chicago: Consortium on Chicago School Research.

Smith, Julia B., Valerie E. Lee, and Fred M. Newmann. 2001 *Instruction and Achievement in Chicago Elementary Schools.* Chicago: Consortium on Chicago School Research.

Smylie, Mark A., Elaine Allensworth, Rebecca C. Greenberg, Rodney Harris, and Stuart Luppescu. 2001. *Teacher Professional Development in Chicago.* Chicago: Consortium on Chicago School Research.

Smylie, Mark, Stacy Wenzel, Elaine Allensworth, Carol Fendt, Sara Hallman, Stuart Luppescu, and Jenny Nagaoka. 2003. *The Chicago Annenberg Challenge: Successes, Failures, and Lessons for the Future.* Chicago: Consortium on Chicago School Research.

*Sporte, Susan E., Macarena Correa, Joe Kahne, and John Q. Easton. 2003. *Chicago High School Redesign Initiative: A Snapshot of the First Year of Implementation.* Chicago: Consortium on Chicago School Research.

Web Sites of Research Organizations Focusing on Chicago School Reform

Consortium on Chicago School Research: www.consortium-chicago.org

Designs for Change: www.designsforchange.org

Center for Urban School Policy at Northwestern University:
www.sesp.northwestern.edu/CUSP/

Colloquium Agenda

Thursday, January 22, 2004

Welcome and Introductions

Robert Ivry
Senior Vice President for Development and External Affairs, MDRC

Goals in Supporting Annual Research Conferences

David Ferrero
Director of Evaluation and Policy Research, Education Programs,
 Bill & Melinda Gates Foundation

Words from the Co-Conveners

Naomi Housman
Coordinator, National High School Alliance
Shirley Schwartz
Director of Special Projects, Council of the Great City Schools

Importance of Rigorous Research

Phoebe Cottingham
Commissioner, National Center for Education Evaluation and Regional Assistance,
 Institute of Education Sciences, U.S. Department of Education

SESSION I: How Far Have We Come and Where Are We Headed?

A Historical Perspective. *Focus:* Issues, both academic and structural, which have been historically significant drivers shaping past and current high school reform initiatives.

Speaker: **Larry Cuban,** Professor Emeritus of Education, Stanford University
Discussant: **Charles Payne,** Sally Dalton Robinson Professor of African-American
             Studies, History, and Sociology, Duke University
Facilitator: **James Kemple,** Senior Fellow, Department of Education,
             Children, and Youth, MDRC

Federal, State, and Local Perspectives. *Focus:* Given the range of stakeholders in high school reform, key concerns that are shaping the perspectives of both policymakers and practitioners at the local, state, and national levels.

> Panelists: **Hans Meeder,** Deputy Assistant Secretary, Office of Vocational and Adult
> Education, U.S. Department of Education
> **Judy Bray,** Consultant, State Education Analyst
> **Shirley Schwartz,** Director of Special Projects, Council of the
> Great City Schools
> Facilitator: **Monica Martinez,** Founder, National High School Alliance;
> Director, Network for the Advancement of Secondary Education

Friday, January 23, 2004

SESSION II: The Role of Research

Is There Any Solid Evidence of Positive Effects for Students? *Focus:* What do we really know about the effects of high school reform initiatives — and how do we know what we know?

> Presenter: **David Stern**, Professor of Education, University of California at Berkeley
> Discussant: **Richard Murnane,** Juliana W. and William Foss Thompson Professor
> of Education and Society, Harvard University Graduate School
> of Education
> Facilitator: **Marsha Silverberg,** Economist, National Center for Education Evaluation
> and Regional Assistance, Institute for Education Sciences,
> U.S. Department of Education

Using Effect Size to Judge Success: How Big Is Big Enough? *Focus:* Effect size is becoming a metric more frequently used to determine outcomes in education research studies. This discussion addresses some of the advantages and challenges of applying this methodology to high school reform initiatives.

> Co-Presenters: **Howard Bloom,** Chief Social Scientist, MDRC
> **Mark Lipsey,** Director, Center for Evaluation Research and Methodology,
> Vanderbilt Institute for Public Policy Studies, Vanderbilt University
> Facilitator: **James Kemple,** MDRC

SESSION III: Studying Approaches to High School Reform

Case Studies. *Focus:* Using a "case study" approach, panelists will discuss research methodology strengths and challenges in researching several types of high school reform initiatives — including school-level interventions, school design principles, and district-level policy levers for change. In addition, a panel of district superintendents will discuss "research questions" that matter to administrators accountable for high school improvement in urban districts.

> Presenter: **Fred Doolittle,** Vice President, Department of Education, Children, and Youth, MDRC

Case Study A: School-Level Models

> Career Academies: **James Kemple,** MDRC
> Talent Development High Schools: **Corinne Herlihy,** Research Associate, Education, Children, and Youth Department, MDRC

Case Study B: School Design Principles

> Gates Schools Initiative: **Barbara Means,** Director, Center for Technology in Learning, Co-Principal Investigator, SRI International
> **David Rhodes,** Senior Research Analyst, American Institutes for Research
> Facilitator: **C. Kent McGuire,** Dean, College of Education, Temple University

What Research Questions Matter to Superintendents?

> Panelists: **Anthony Amato,** Superintendent, New Orleans Public Schools
> **Bernard Taylor,** Superintendent, Kansas City, Missouri, School District
> Facilitator: **Janet Quint,** Research Manager, First Things First Evaluation**;** Senior Associate, Department of Education, Children, and Youth, MDRC

Case Study C: A District Strategy for High School Reform

> Presenter: **G. Alfred Hess,** Research Professor of Education and Social Policy, Northwestern University, Center for Urban School Policy
> Discussant: **George Bohrnstedt,** Senior Vice President for Research, American Institutes for Research
> Facilitator: **C. Kent McGuire,** Temple University

Reflections on the Colloquium

Key Themes: **Glee Holton,** Director of Development, MDRC
Panelists: **Phoebe Cottingham,** Institute of Education Sciences
                    David Ferrero, Bill & Melinda Gates Foundation
                    Steve Fleischman, Principal Research Scientist, American Institutes
                        for Research
Facilitator: **Fred Doolittle,** MDRC

Closing Remarks

Robert Ivry, MDRC

Presenters and Panelists

Anthony Amato is Superintendent of New Orleans Public Schools. Previously, he served as Superintendent of Hartford Public Schools, where he implemented a reorganization of the school district and instituted a series of programs and curriculum reforms aimed at bringing stability and academic excellence to Connecticut's largest school district. Prior to his work in Hartford, Amato spent 12 years as Superintendent of New York City School District 6, which he helped raise from being the city's lowest-performing district to being rated fifteenth out of 32 districts citywide.

Dr. Howard Bloom is Chief Social Scientist for MDRC, where he leads the development of experimental and quasi-experimental methods for estimating program impacts and also works closely with MDRC staff to build these methods into research designs. Prior to joining MDRC, he taught research methods, program evaluation, and applied statistics for 21 years at Harvard University and at New York University.

Dr. George W. Bohrnstedt is Senior Vice President for Research at the American Institutes for Research (AIR), where he is involved in the development of new programs of research, especially in the area of education. He also leads AIR's Council of Chief Scientists in the development of new applied methodological and statistical tools. Bohrnstedt has a deep interest in issues of education research and policy at the K-12 level. He currently chairs the National Assessment of Educational Progress Validity Studies Panel at the National Center for Education Statistics. Formerly, he was the principal investigator of the Bill & Melinda Gates Foundation-funded evaluation of an initiative to create small, personalized high schools and was the Co-Principal Investigator of the evaluation of California's K-3 Class Size Reduction Program.

Judy Bray is a state education analyst with a national focus. With more than two decades of experience in the state policy arena, she helps state leaders cut through the intricacies of education policy and practice. Bray writes on topics ranging from accountability to service learning, presents policy analysis in a wide array of settings, and facilitates state work groups who wish to take action. Her consulting practice supports educators and policy leaders working together to analyze, initiate, and sustain system reform efforts.

Dr. Phoebe Cottingham is Commissioner of the National Center for Education Evaluation and Regional Assistance at the U.S. Department of Education's Institute of Education Sciences (IES). Before joining IES, Cottingham was the senior program officer for domestic public policy at the Smith Richardson Foundation, where she developed priorities and strategies to fund innovative projects on school reform and early childhood education. Previously, Cottingham served as Associate Director of The Rockefeller Foundation's Equal Opportunity Program,

where she oversaw projects dealing with minority single parenting, community-based employment programs, and child care policy.

Dr. Larry Cuban is Professor Emeritus of Education at Stanford University. He taught social studies for fourteen years in inner-city high schools and worked for seven years as a district superintendent. Cuban also worked directly with Bay Area teachers and administrators as a sponsor of a social studies alumni group of the Stanford Teacher Education Program (STEP). Trained as a historian, Cuban has written extensively about the history of school reform, leadership, teaching, and the uses of technology in schools.

Dr. Fred Doolittle is Vice President of MDRC's Department of Education, Children, and Youth, which specializes in studies of programs for economically disadvantaged youth. He is currently directing MDRC's evaluation of the Scaling Up First Things First Initiative and is the Research Director for the National Evaluation of Project Graduation Really Achieves Dreams (Project GRAD). Doolittle is also overseeing a multiyear demonstration and evaluation of reading and math curricula in after-school programs throughout the country for the U.S. Department of Education.

Dr. David Ferrero has worked as a journalist, high school teacher, policy researcher, educational technology consultant, and state policy consultant on teacher quality. He has master's degrees in English and American studies and a doctorate in education policy from Harvard University. As Director of Evaluation and Policy Research for education programs at the Bill & Melinda Gates Foundation, Ferrero oversees policy initiatives and program evaluation, and he co-manages the foundation's national high school program.

Steve Fleischman is Principal Research Scientist at the American Institutes for Research, where he specializes in the identification and successful implementation of effective education programs and practices. With nearly 20 years of education experience, Fleischman has served for the past 10 years as a director of and advisor to many education projects, including "Educators' Guide to Schoolwide Reform," "Guidelines for Ensuring the Quality of National Design-Based Assistance Providers," and "Standards for Web-Based Education Products and Services."

Corinne Herlihy is Research Associate at MDRC and currently serves on the Talent Development Evaluation team, for which she has developed data management strategies for quantitative data obtained from school districts. She also conducts analyses of school- and student-level data for Talent Development. Herlihy co-authored MDRC's report commissioned by the Council of the Great City Schools, *Foundations for Success,* a compilation of case studies of promising district-wide efforts to raise achievement and close achievement gaps in underserved communities. Prior to her graduate studies at Harvard University, where she earned a master's degree in public policy at the John F. Kennedy School of Government, Herlihy was a teacher of mathematics at the middle and high school levels.

Dr. G. Alfred Hess, Jr., is Research Professor of Education and Social Policy, Director of the Center for Urban School Policy, and Coordinator of The Lighthouse Project at Northwestern University, where he studies educational policy, urban public education, and school finance. Previously, he directed the Chicago Panel on School Policy, and, in this capacity, he was one of the founders of the Chicago School Reform Act of 1988. Currently, he directs evaluation projects for the Chicago Public Schools and coordinates Northwestern University's partnership with Evanston School District 65.

Glee Ivory Holton is Director of Development at MDRC. She has served as a senior manager for education programs and has had extensive experience working with youth programs, schools, and school districts. Holton was a court-appointed monitor for the Cleveland School Desegregation Case. She is currently Director of the Project GRAD national evaluation.

Naomi Housman is Coordinator for the National High School Alliance, based at the Institute for Educational Leadership in Washington, DC. She has focused her career on serving public school systems in high-poverty urban communities, with a particular emphasis on secondary education. In her current role, Housman helps to mobilize the resources, knowledge, and capacity of the partner organizations to work collectively in shaping policy, research, practice, and public engagement. The vision of the National High School Alliance is to foster high achievement, close the achievement gap, and promote civic and personal growth among all youth in our high schools and communities.

Robert J. Ivry is Senior Vice President for Development and External Affairs at MDRC. He is a nationally known expert on social policy issues, especially in the areas of workforce development, school-to-career, education reform, welfare reform, and youth development/employment. Ivry plays a major leadership role at MDRC, working with the federal government, states, localities, school districts, and community-based organizations to develop and test innovative ideas designed to improve the economic well-being and life circumstances of low-income populations. Currently, he directs a demonstration called Opening Doors, aimed at helping low-income individuals earn college credentials, and he coordinates MDRC's work with at-risk youth.

Dr. James J. Kemple is Senior Fellow in MDRC's Department of Education, Children, and Youth, also serving as a senior advisor and policy analyst for the department. He has served as principal investigator and research director on a variety of MDRC's education, employment and training, and welfare-to-work program evaluations. He is the principal investigator for the Career Academies Evaluation and the National Evaluation of the Talent Development Model. He is heading the design and site recruitment tasks for the National Reading First Impact Study. As a practitioner, Kemple taught high school math and managed a three-phased supplementary academic and high school placement program for disadvantaged youth in Washington, DC.

Dr. Mark Lipsey is Director of the Center for Evaluation Research and Methodology and a Senior Research Associate at the Vanderbilt Institute for Public Policy Studies at Vanderbilt University. He received a doctorate in psychology from Johns Hopkins University, following a bachelor's degree in applied psychology from Georgia Institute of Technology. His recent research has mainly involved the application of meta-analysis techniques to identify effective intervention programs and predictive risk factors for juvenile antisocial behavior, issues of methodological quality in program evaluation research, and ways to apply research findings to improve program practice. He has published books, articles, and technical reports in these areas, including *Practical Meta-Analysis* (with David Wilson) and *Evaluation: A Systematic Approach,* 7th Edition (with Peter Rossi and Howard Freeman).

Dr. Monica Martinez is Director of the Network for the Advancement of Secondary Education, a new center at the Institute for Education Leadership (IEL) in Washington, DC. She serves as the Project Director for IEL's work with the National Clearinghouse for School Reform, the Theme High Schools Network, and the Catalog of Research on Secondary School Reform. She is the founder of and senior advisor to the National High School Alliance, and she oversees IEL's work with the Pathways to College Network. She has worked in a variety of higher education institutions and intermediary organizations that provide programmatic assistance in partnership development, school change, and research and evaluation. Her work has focused on issues related to educational access and achievement for low-income and minority students.

Dr. C. Kent McGuire is Dean of Temple University's College of Education. Previously, he was Senior Vice President of MDRC, where his responsibilities included leadership of the Department of Education, Children, and Youth. From 1998 to 2001, McGuire served in the Clinton administration as Assistant Secretary of the U.S. Department of Education, where he was the senior officer for the department's research and development agency. As Education Program Officer for the Pew Charitable Trusts from 1995 to 1998, McGuire managed Pew's K-12 grants portfolio. From 1991 to 1995, he served as Education Program Director for the Eli Lilly Endowment.

Dr. Barbara Means directs SRI International's Center for Technology in Learning and serves as co-principal investigator of the AIR/SRI national evaluation of the Bill & Melinda Gates Foundation National School District and Network Grants Program. Her research focuses on ways to foster students' learning of advanced skills and the changes in practice at the school and classroom levels associated with the introduction of technology-supported innovations. Means served on the National Academy of Sciences' Committee on Developments in the Science of Learning, which produced the volume *How People Learn*, and as a member of the Academy's Board on Testing and Assessment (BOTA). Her published works include the edited volumes

Teaching Advanced Skills to At-Risk Students, Evaluating Educational Technology, and *Technology and Education Reform.*

Hans Meeder is Deputy Assistant Secretary of the U.S. Department of Education's Office of Vocational and Adult Education, where he is responsible for directing research and dissemination activities in support of career and technical education in high schools and colleges, adult basic education, and English language acquisition. His background in education public policy includes a broad expertise in workforce trends, research on effective practice, and education accountability systems.

Dr. Richard Murnane is Juliana W. and William Foss Thompson Professor of Education and Society at Harvard University's Graduate School of Education, where he focuses his research on the relationships between education and the economy, teacher labor markets, the determinants of children's achievement, and strategies for making schools more effective. In his book *Who Will Teach? Policies That Matter* (with J. Singer and J. Willett), Murnane showed that teachers' salaries and certification requirements strongly affect the composition of the public school teaching force. Murnane's *Teaching the New Basic Skills*, coauthored with Professor Frank Levy of Massachusetts Institute of Technology, explains how changes in the U.S. economy have increased the number of skills that high school graduates need in order to earn a middle-class living and shows how schools must change to provide all students with the requisite skills. Murnane serves on MDRC's Board of Directors.

Dr. Charles Payne is Sally Dalton Robinson Professor of African-American Studies, History, and Sociology at Duke University and Faculty Affiliate at the Center for Child and Family Policy. His research focuses on urban education, the civil rights movement, social change, and social inequality. Payne has been named a 2004 Carnegie Scholar by the Carnegie Corporation. He is also the recipient of the Spencer Foundation's Senior Scholar Award, a four-year grant that he is using to develop a book about the persistence of failure in urban schools and what we have learned in the past decade or so about creating large-scale improvement in urban districts. Payne serves on MDRC's Board of Directors.

Dr. Janet Quint is Senior Research Associate at MDRC, where she currently is the Research Manager for the Scaling Up First Things First evaluation and lead author of the project's two reports to date. She was the director of Project Transition, a demonstration focused on the transition of young people to high school. Quint is one of MDRC's leading implementation researchers. A sociologist by training, she also holds a Master of Arts in Teaching.

Dr. David Rhodes is Senior Research Analyst at the American Institutes for Research, where he is the principal investigator for a project supporting the U.S. Department of Education's Quality Assurance Program and coordinator of quantitative data analysis in the evaluation of the Bill & Melinda Gates Foundation's support of smaller learning communities within our nation's

high schools. Rhodes has overseen the development and implementation of numerous methodological research designs and authored reports on a broad range of topics, including the relationships between school and teacher characteristics and student achievement, quality assurance in financial-aid delivery, the social policy implications of changes in the price of higher education, and discrimination in the awarding of public contracts.

Dr. Shirley Schwartz is Director of Special Projects at the Council of the Great City Schools, where she oversees programs and initiatives that focus on improving teaching and learning in urban schools, including several major projects to recruit and prepare a diverse and highly qualified teacher workforce. Schwartz also serves as the liaison to the council's affiliate, the Council of the Great City Colleges of Education, and is a member of several editorial and national advisory boards that focus on urban teacher preparation and quality. Before joining the Council of the Great City Schools, Schwartz was Associate Dean of the School of Professional Studies in Washington, DC, and a Research Associate at the University of Maryland's Institute for the Study of Exceptional Children and Youth.

Heather Schweder is a consultant to MDRC and works as a conference planner, researcher, and writer for education and communication projects.

Dr. Marsha Silverberg, an economist, is a team leader at the U.S. Department of Education's Institute of Education Sciences (IES), National Center for Education Evaluation and Regional Assistance. She designs and oversees rigorous impact evaluations of high school improvement and school choice strategies. She is currently concluding her assignment as Director of the National Assessment of Vocational Education (NAVE) for the department, having prepared a mandated report to Congress on the implementation and effects of vocational education and federal vocational legislation. Prior to joining the Department of Education, she was a senior researcher at a policy research firm, where she led numerous studies of youth interventions, particularly those intended to improve students' transitions from school to work.

Dr. David Stern is Professor of Education at the University of California at Berkeley, where he teaches economics of education and conducts research on high school reform and the relationship between education and work. From 1995 to 1999, Stern served as Director of the National Center for Research in Vocational Education, based at Berkeley's Graduate School of Education. From 1993 to 1995, he was principal administrator in the Center for Educational Research and Innovation at the Organization for Economic Cooperation and Development in Paris.

Dr. Bernard Taylor, Jr., is Superintendent of the Kansas City, Missouri, School District. He joined the district in 2000 as Executive Director for School Leadership, supervising the principals and their staff at 26 elementary schools as well as at King Middle School and Southeast High School.

About MDRC

MDRC is a nonprofit, nonpartisan social policy research organization. We are dedicated to learning what works to improve the well-being of low-income people. Through our research and the active communication of our findings, we seek to enhance the effectiveness of social policies and programs. MDRC was founded in 1974 and is located in New York City and Oakland, California.

MDRC's current projects focus on welfare and economic security, education, and employment and community initiatives. Complementing our evaluations of a wide range of welfare reforms are new studies of supports for the working poor and emerging analyses of how programs affect children's development and their families' well-being. In the field of education, we are testing reforms aimed at improving the performance of public schools, especially in urban areas. Finally, our community projects are using innovative approaches to increase employment in low-income neighborhoods.

Our projects are a mix of demonstrations — field tests of promising program models — and evaluations of government and community initiatives, and we employ a wide range of methods to determine a program's effects, including large-scale studies, surveys, case studies, and ethnographies of individuals and families. We share the findings and lessons from our work — including best practices for program operators — with a broad audience within the policy and practitioner community, as well as the general public and the media.

Over the past quarter century, MDRC has worked in almost every state, all of the nation's largest cities, and Canada. We conduct our projects in partnership with state and local governments, the federal government, public school systems, community organizations, and numerous private philanthropies.