

PACE

POLICY ANALYSIS FOR CALIFORNIA EDUCATION

**Looking Back: Program
Successes and Evaluations Under
"Jobs for the Disadvantaged"**

Charles Dayton

March 1988

Directors

**James W. Guthrie
University of California at Berkeley**

**Michael W. Kirst
Stanford University**

**Allan Odden
University of Southern California**

POLICY ANALYSIS FOR CALIFORNIA EDUCATION

**Looking Back: Program
Successes and Evaluations Under
"Jobs for the Disadvantaged"**

Charles Dayton

March 1988

Charles Dayton is a policy analyst with PACE.

This paper was sponsored and published by Policy Analysis for California Education, PACE. PACE is funded by the William and Flora Hewlett Foundation and directed jointly by James W. Guthrie and Michael W. Kirst. The analyses and conclusions in this paper are those of the author and are not necessarily endorsed by the Hewlett Foundation.

Executive Summary

In early 1980, the Clark Foundation launched an ambitious series of demonstration programs designed to address the high rate of school dropouts and youth unemployment in several American cities. These programs shared a focus on disadvantaged minority youth, but they varied in their structure from site to site—from a focus on job search and placement in grades 11 and 12, to academic skills and vocational training throughout high school.

Beginning with the 1984-85 school year, the evaluations' emphasis moved from technical assistance and process evaluation to assessing changes in student outcomes. A matched comparison group design was used to determine whether program students were making gains in attendance, credits earned, grades, and standardized test scores. In only one site, Chicago, were statistically significant differences consistently found between program and comparison sites on these measures. The three academy sites (Palo Alto, Pittsburgh, and Portland) showed some limited evidence of effects in these realms, while the remaining sites did not. All sites showed gains in school retention and student attitude measures.

Another Clark Foundation objective was to influence the institutions in each of the cities to be more responsive to disadvantaged minority youth. The demonstration programs have shown considerable success in this sense. The Boston Compact has become a national model. There are statewide replications underway of the Clark Foundation-funded programs in California and Colorado. In many sites, the programs have had an impact on the cities and may be replicated at that level.

This work has led to a number of lessons and insights about conducting such evaluations. The many issues one must confront are reviewed both in designing such evaluations and in obtaining necessary data from schools. In addition, the purposes such evaluations serve, and guidelines to be followed in conducting them, are also reviewed.

Contents

Executive Summary	iii
Foreword	vii
Policy Analysis for California Education	ix
Introduction	1
The Clark Foundation's Mission.....	1
The Changing National Context.....	1
The Programs and Their Performance	2
Differing Program Models.....	2
Importance of Program Contexts	4
Student Outcome Findings.....	5
Process Findings.....	6
Political Impact.....	8
Evaluation Lessons	9
Alternative Evaluation Designs.....	9
Evaluation Problems.....	11
Evaluation Guidelines	13
Conclusions	14
Appendix A: SWAP and Academy Program Treatments	15
Appendix B: Program/Comparison Group Differences, 1984-1987	17
Appendix C: Political Impact	19

Foreword

In the spring of 1981, the Edna McConnell Clark Foundation hired the American Institutes for Research (AIR) to provide technical and evaluation assistance to its Jobs for the Disadvantaged Program and its four initial demonstration sites: Akron, Albuquerque, Boston, and Philadelphia. As a research scientist in AIR's Palo Alto office who was interested in youth employment, I became the evaluation site manager for the Albuquerque program.

During the past seven years I have worked with four Clark Foundation program directors, Myrtis Mosley, Hattie Harlow, Robyn Govan, and Hayes Mizell, and with 13 School-to-Work Action (SWAP) and Academy programs. For the past three years I have been responsible for conducting the student outcomes evaluation of these programs, as well as the process evaluation this past year.

The evaluation has gone through a considerable transition since its beginnings in 1981. From its original focus on technical assistance, it added process evaluation, for which information was collected on how well the programs were being implemented. By the 1984-85 school year, the evaluation focus shifted toward assessing student outcomes, and this emphasis has continued since.

As both the Jobs for the Disadvantaged Program and my work as evaluator moved toward an end, I felt it was appropriate to tie together the knowledge and insights gained during the past seven years. What did the Clark Foundation hope to accomplish? How well did the funded sites perform and why? What were the issues encountered in evaluating the programs? When I approached Hayes Mizell with this idea, he supported it and encouraged me to make this attempt.

It is my hope in this summary analysis to bring together the knowledge and insights developed over these years in a succinct form. I have appreciated the opportunity to conduct this work and thank the many others who have helped in this endeavor. Foremost among these are the Clark Foundation program officers mentioned above, Dr. Victor Rouse, who directed the evaluation efforts through 1984-85, and Dr. Alan Weisberg, who has worked with me over the past year. Acknowledging this help, I take full responsibility for the conclusions presented here.

Charles Dayton

Policy Analysis for California Education

Policy Analysis for California Education, PACE, is a university-based research center focusing on issues of state educational policy and practice. PACE is located in the Schools of Education at the University of California, Berkeley and Stanford University. It is funded by the William and Flora Hewlett Foundation and directed jointly by James W. Guthrie and Michael W. Kirst. PACE operates satellite centers in Sacramento and Southern California. These are directed by Gerald C. Hayward (Sacramento) and Allan R. Odden (University of Southern California).

PACE efforts center on five tasks: (1) collecting and distributing objective information about the conditions of education in California, (2) analyzing state educational policy issues and the policy environment, (3) evaluating school reforms and state educational practices, (4) providing technical support to policy makers, and (5) facilitating discussion of educational issues.

The PACE research agenda is developed in consultation with public officials and staff. In this way, PACE endeavors to address policy issues of immediate concern and to fill the short-term needs of decision makers for information and analysis.

PACE publications include Policy Papers, which report research findings; the Policy Forum, which presents views of notable individuals; and Update, an annotated list of all PACE papers completed and in progress.

Advisory Board

Mario Camara
Partner
Cox, Castle & Nicholson

Constance Carroll
President, Saddleback
Community College

Gerald Foster
Region Vice President
Pacific Bell

Robert Maynard
Editor and President
The Oakland Tribune

A. Alan Post
California Legislative Analyst,
Retired

Sharon Schuster
Executive Vice President
American Association of University Women

Eugene Webb
Professor, Graduate School of Business
Stanford University

Aaron Wildavsky
Professor of Political Science
University of California, Berkeley

Introduction

The Clark Foundation's Mission

In 1980, youth unemployment was an important issue on the national agenda. The unemployment rate for youth, and particularly minority youth, was at a historic high. The baby boom generation had been entering the labor market in record numbers for a decade. In the late 1970s, President Carter and Congress had focused considerable energy and money on national efforts to address the needs of such youth. These efforts led to sizable employment programs, both private and public, as well as to a number of new training approaches, and the information from these programs and experiments was just emerging.

The Clark Foundation entered this picture concerned that the core of the unemployment problem, inner-city minority youth, was not being served adequately by federal programs. It recognized the relationship between youth unemployment and other problems, such as drugs, crime, and inadequate education. It was also concerned that the private sector was not playing a strong role in the new efforts.

The Clark Foundation chose to take a preventive approach and focus on youth still in school rather than those who had dropped out. It hoped to select a few sites and fund well designed school-based programs tied to partnerships with the business and larger communities in selected cities. As a result, it hoped to have a significant impact on institutions in those cities and on the fortunes of at-risk youth.

While the problem was clear and pressing, how best to proceed was less clear. Knowledgeable people were involved at most sites and were given broad scope in designing their programs. They were encouraged to define their own local solutions and, in fact, developed rather different approaches. These ranged from short-term job search and placement programs to in-depth efforts to improve academic skills.

The Changing National Context

Meanwhile, conditions in the nation were changing. President Reagan's election refocused attention and priorities. Unemployment was downgraded as a concern, seen primarily as an economic issue; the economy entered a period of strong recovery. Under the Job Training Partnership Act (which replaced CETA), Private Industry Councils grew in importance, resulting in much greater influence on job training efforts by the private

sector and less focus on the truly at-risk. Federal dollars dwindled. Baby boom labor market entrants declined, resulting in declines in overall youth unemployment statistics.

The problem with this scenario is that viewing youth unemployment as primarily an economic issue ignores the distinction between cyclical and structural unemployment. There are many disadvantaged young people who lack the skills to find suitable employment in the best of economies. During the past few years, teenage unemployment has stayed at about 2.5 times the level of overall unemployment and has declined as the overall rate has declined. Unemployment among disadvantaged youth has changed little. In 1986, when overall unemployment averaged 7.0 percent (18.4 percent for teenagers), the rate for black teenagers was 39.3 percent. In short, while national consensus about the importance of urban youth unemployment dissipated, the problem continued largely unabated.

THE PROGRAMS AND THEIR PERFORMANCE

Differing Program Models

The Clark Foundation's selection of sites for its "School-to-Work Action" programs (a name used generically initially) has consistently focused on the intended target group, which is urban and minority. Indeed, 95 percent of the students participating in these programs during the 1986-87 school year were black or Hispanic. But the programs themselves have varied in their structure from the beginning.

Among the first round of funded programs, Boston represented one extreme. The Boston Compact offered some brief job search and placement help in grades 11 or 12, initially in a few but eventually in all of Boston's high schools. While the Compact's "treatment," as provided through the Jobs Collaborative, was thin, the Compact brought the attention of the whole city to the problem and resulted in large numbers of job placements for both in-school students and graduates.

At the other extreme, in Philadelphia, the Clark Foundation supported a SWAP program with a much fuller treatment, over three years, in grades 10-12, focused primarily on academic remediation. It operated in just one high school, had low visibility, and developed an unfortunate "remedial" stigma. Eventually it became a dumping ground for both teachers and students and had little measurable impact on either the students involved or the city.

The Albuquerque "Career Guidance Institute" focused on career awareness, building school-business partnerships, in part based on the Adopt-a-School model. Working through the local chamber of commerce, and operating at first in one high school and its two feeder middle schools and eventually throughout the city, it sponsored field trips, business speakers, and career days for students; made summer job placements, and provided career-related staff development for teachers.

The last of the original programs was to be in Akron. But the program there never got off the ground, and the foundation soon withdrew its support.

Slightly later, in 1981, the Peninsula Academies in Palo Alto received a three-year grant. The Academies, located in two high schools in the Sequoia School District just north of Palo Alto, incorporates students from East Palo Alto who attend high school in this district. It combined technical training in computers and electronics with a school-within-a-school structure of academic courses. The program, which operates in grades 10-12, enjoyed good corporate support, as business came through with both mentors and jobs.

By the 1984-85 school year, six new programs were added to the original group: Chicago, Denver, East St. Louis, Pittsburgh, Portland, and Washington, D.C. Two of these were modeled on the Peninsula Academies: Pittsburgh and Portland. Three were hybrids of the Academy and Compact models: Chicago, Denver, and East St. Louis. Washington, D.C. settled on a new approach, focusing on staff development designed to train teachers in how to better provide school-to-work transition help to students. This led to an extensive curriculum writing effort and eventually to an Academy-like approach for students.

Finally, beginning in the 1985-86 school year, Cleveland and Oakland were added. Cleveland is modeled on the Washington program and is focused on staff development leading to curriculum development. Oakland is modeled on the Boston Compact, providing brief job search training and job placements.

To try to illustrate the wide variety of "treatments" among these programs, during the 1985-86 school year I developed a system for estimating the amount of time students spent in each program in each of five types of activities. These were:

- employability skills/job preparation
- vocational training
- academic classes
- enrichment activities
- work experience

I found that actual contact hours by students varied greatly, from a few hours per week for a few weeks at some sites, to more than 20 hours per week over three or four years at others.

To illustrate these differences, I have presented in Appendix A a table that summarizes the number of hours a given student spends in each of the above categories of activity over the course of the program. This table illustrates the wide variations in program treatments. At one end of the spectrum are programs based on the Boston Compact model, which consist primarily of some brief job search assistance coupled with extensive work experience in grades 11 or 12. At the other end are the Academies in Palo Alto, Pittsburgh, and Portland, the Chicago Job Readiness Program, and the Public-Private Partnership Program that evolved in Washington, D.C. Each of these has multi-year treatments with substantial academic and vocational elements.

Importance of Program Contexts

In addition to the differences in how the various programs are structured, they also operate in very different settings. Although all sites chosen were in urban areas, the quality of the public school systems, health of the local economy, and nature of the youth population served varied considerably.

Two contrasting sites, East St. Louis and Palo Alto, provide an illustration. Average family income in East St. Louis is among the lowest in the country; 73 percent of that district's students come from families below the poverty line. Average family income in San Mateo County, in which the Sequoia School District and Peninsula Academies operate, was \$55,000 in 1985. East St. Louis test scores are two-and-a-half years below the national norm, while those of the Sequoia School District are among the highest in California. The unemployment rate in East St. Louis hovers around 20 percent, while in the Silicon Valley near Palo Alto it is under 5 percent and entry-level jobs often go begging. The East St. Louis school population is nearly 100 percent black, while the minority population of the Sequoia School District is about 25 percent. It is very difficult to judge on the same terms programs in such widely differing settings.

These differences in setting and context interacted with the differing program models in each site to produce widely differing results. In Boston, a powerful political and business community came together in a way that would have ensured some impact regardless of the program model. In Chicago, a director was chosen who most likely would have made something happen in almost any setting. In Palo Alto, the richness of the environment increased the chances of success. By contrast, the severity of both school and community problems in North Philadelphia and East St. Louis severely reduced their chances of success. Labor markets like those in East St. Louis, Oakland, and Pittsburgh

made development of jobs difficult. It is important to understand these environmental influences on the programs in making judgments about their success.

Student Outcome Findings

How should the programs be judged? What should be the criteria used in the evaluations? Ideally a program should be judged in terms of its objectives. If you ask the persons who design and operate the various Clark Foundation programs what their main objectives are, the answers are usually something like "To give at-risk youth a reason to keep trying and to graduate" or "To give disadvantaged kids a better chance to make it." Even the foundation's objectives for the programs were at the level of statements like "demonstrating successful models" or "institutionalizing a process of change in the schools." These were worthwhile goals, but they are not easy to measure.

As the Clark Foundation moved toward more rigorous evaluation, it became necessary to translate the programs' goals into measurable "indicators" and collectible data. This adds numerical precision; but it also substituted proxies for what the programs said they wanted to accomplish. In social science research there often is a simple inverse relationship between what is measurable and what is important.

Since changes in student behavior are central to the programs' goals, the measurable indicators arrived at were primarily related to academic performance—attendance, credits, grades, and standardized test scores. Some attitudinal indicators were obtained through pre-post program student questionnaires. And perhaps most central, retention in school was tracked. Matched comparison groups—students like those in the program in terms of age, gender, ethnicity, and past school performance—were identified and tracked on these same academic measures. The year-to-year evaluations thus tell us whether program students are outperforming their nonprogram peers on these measures.

What did the programs accomplish in terms of student outcomes? The last three years' evaluation reports provide considerable information related to this question. Some readers have complained that there has been too much information in these reports. They have asked for simplifications and judgments about what all the data mean. In the table in Appendix B, I have attempted this, providing a site-by-site summary of student outcome results over the past three years.

As the table shows, in terms of statistically significant differences between program and comparison groups, the results of the last three years are quite mixed. Chicago has consistent evidence of positive effects. There is evidence in the three Academies: Palo Alto, Pittsburgh, and Portland. There is little or no evidence of such effects elsewhere. It should be understood that it is relatively rare in such student outcomes oriented evaluations

to find clear examples of success; most educational evaluations turn up the finding of "no significant differences" between treatment and nontreatment groups. Thus even limited evidence of impact at a statistically significant level may be regarded as encouraging. On the other hand, if one spends considerable sums on a program, one seems entitled to expect substantial results.

Not all program effects are represented in this table. For example, almost all the programs have shown some effect in terms of reduced dropouts, and this was particularly true in 1986-87. The problem with dropout data is that they are relatively unreliable, and they are handled differently from site to site. They are also less open to meaningful statistical tests since they are categorical in nature and do not offer a scaled score. Nevertheless, they are an important indicator for these programs, and all sites where they were collected in 1986-87 showed positive effects.

The pre-post student questionnaires show certain types of fairly consistent positive changes over time as well, and again this was particularly true in 1986-87, when changes over three years could be observed. Program participants at all sites report substantial increases in career-related experiences, and most show advancements in their career-related plans and attitudes. Some report more positive feelings toward school and themselves, although these changes have typically been small. And in all sites students see improved career opportunities as a result of the program and report positive feelings toward the program. Most sites also have predominantly positive staff feedback.

Process Findings

The student outcome evaluations address the question of *what* has been accomplished, but not *how* and *why*. This is the province of the process evaluation, which examined how well the programs were implemented and what factors seemed to determine their effectiveness. By examining variations in settings, models, and quality of implementation from site to site, one can begin to arrive at conclusions about what is required for success.

The process evaluations have identified a variety of issues related to program implementation. In reviewing these reports one finds commonalities from year to year and from site to site. Given problems or strategies led to similar outcomes time after time. A summary of these factors that have played an important role across the sites for the past several years provides a set of guidelines regarding the factors that lead to program success.

Since no one factor operates in the absence of the others or is a determinant entirely by itself, perhaps the best way of stating these is as a series of necessary but not sufficient conditions:

Setting

1. Although difficult settings are to be assumed, the setting must not be so deprived in either educational quality or labor market health as to preclude success
2. Support at high levels in the host educational and business communities
3. Sufficient flexibility within the district, high school, and supporting companies, to permit the variations in structure and schedule required by the program

Program Model

4. Clearly defined and realistic objectives
5. A sufficiently substantive treatment that if well implemented it can reasonably expect to influence students in the desired ways
6. A clearly defined, sensible, and consistent student selection procedure

Implementation

7. Sufficient time allowed for the program to overcome inevitable startup problems and establish itself
8. Strong personnel: well-organized managers, effective teachers, and a sufficient supporting cast (administrators, counselors, parents, employees, community members)
9. Sufficient program resources: funding, facilities, equipment, and supplies
10. An evaluation/management feedback system that leads to program refinements and, where possible, provides evidence of success when it occurs

Of course most of these characteristics are not simply present or absent but present to some degree. Probably no program has them all to the degree that would be ideal. And they interact with each other. A strong model will drive the achievement of many of these. Strong managers will find ways of making them happen. Sufficient resources and a good student selection procedure will lead to a positive program identity. The more of these that are present, the better the chance for success. The more that are lacking, especially to a serious degree, the greater the certainty of failure.

Perhaps some examples will illustrate. Chicago is an interesting example because while the program operates in a very difficult setting it nevertheless succeeds. The Job Readiness program here uses a substantive model, with all five types of treatment activities discussed earlier, over several years. It is directed by a strong leader who insists on high quality teachers and staff and who finds ways of obtaining needed resources. She also builds both school and corporate support through the confidence she engenders. Initial evaluation findings were positive. The program developed a positive identity, easing its student selection and resource development. It has become a cycle: good management has led to strong staffing and success which has brought the resources and recognition necessary to maintain that success.

Philadelphia offers a contrasting example. The SWAP there also operated in a difficult setting, Simon Gratz High School in North Philadelphia. Its program model was strong academically but weak in terms of private sector support and activities. It lacked leadership; consequently it failed to develop a corps of strong teachers, who were also hurt by the lack of other resources that failed to develop. This became a negative cycle: teachers were unhappy, turnovers were common, and there was no esprit de corps; students were unhappy, especially at the lack of business experiences, making recruitment difficult; initial negative evaluation results furthered the negative image. Eventually the program developed a stigma and was terminated with a clear sense of failure.

Between these two extremes there are many other examples of success or failure related to one or another of the necessary conditions for success. What these examples demonstrate is that if the right mixture of setting, program model, and implementation come together, success will follow. But there are many reasons why programs can fail, and only if a whole series of conditions is met to at least a reasonable degree will they succeed.

Political Impact

The Clark Foundation has made clear from the start that it has two objectives in each site: (1) to establish a successful program and (2) to advance the cause of at-risk

youth by influencing institutions to recognize and respond to their needs. This second objective has been less directly measured by the evaluations, but it should not be ignored. It can be assessed in a number of ways:

- The support given the program by the school and district
- The support given the program by the private sector
- Whether the program continues to operate after foundation funding expires
- The attention the program garners in the city and beyond
- Whether the program is replicated within the city, state, or nation

Thus in addition to the assessments of the programs' impact on their participants, I have made judgments about each site's success in this "political" sense. These are presented in the chart in Appendix C.

As this chart shows, there is a mixed picture of success among the 13 programs. On balance, they have had more success in the political sense than in improving student outcomes. This is particularly true in Boston, where student outcome progress could not be measured. But the Compact model has achieved national prominence. It is also true in Denver, where the program *is* evaluable and has had little or no measurable effects, but where there are nonetheless 11 replications of the program underway throughout Colorado with more planned for next year.

In California, the Peninsula Academies have evolved into a statewide model with over \$1 million expended annually on replications by the state. In Pittsburgh, Portland, and Washington the programs have had impact as models within the city and either have been or may be replicated locally. These are notable accomplishments. What they suggest is that success in terms of student outcomes is not invariably necessary to effect institutional change.

EVALUATION LESSONS

Alternative Evaluation Designs

The Clark Foundation has expended a considerable amount on the evaluations of these programs over the past seven years. While the central question regards what has

been learned about the programs, also of interest is what has been learned about evaluating them.

There are essentially two forms of evaluation: process and outcome. The first examines the implementation of programs and provides feedback to managers in order to help them refine their efforts. It is based largely on observation and interviews. The second, outcome evaluation, examines changes in student performance and provides evidence of impact on students, usually of interest to funding agencies as well as program managers. Outcome evaluations rely on more structured data collection and rely on statistical tests to draw their conclusions.

Most evaluators believe both types of evaluation are important and that they interact with each other. Feedback to managers regarding program implementation is of limited value if one does not know whether the program has made any difference to students. On the other hand, knowing that a program has had a substantial influence on student performance is interesting, but knowing why is far more valuable. As discussed earlier, since the early 1980s the Clark Foundation-sponsored evaluations have swung from one end of the spectrum to the other. They finally settled in the middle, with both process and outcome elements included in the final 1986-87 work.

The student outcomes evaluation design employed in these evaluations uses a matched comparison group, which is a quasi-experimental design. The chief alternative is a true experimental design in which students from one large pool are randomly assigned either to a program or a nontreatment "control" group. The randomization ensures a good match between the two groups, whereas with the matched comparison group design, one must try to match the program students on a post-hoc basis. This requires obtaining information on both program and nonprogram students regarding matching variables, such as gender, ethnicity, and pre-program school performance. One can never control for everything, and obtaining all this data is both laborious and uncertain, so it still leaves inevitable questions. In short, it requires more work to achieve a less certain match.

The advantage of the comparison group design is that the students most appropriate for the program are enrolled in it, based on a human selection process which incorporates the judgment of teachers and counselors and the interest of students. Random assignment is usually resented by both school staff and students and can lead to subversion of the evaluation and even the program. Ultimately one has to decide whether serving students or conducting research is more important; the two are not fully compatible.

A third alternative is a single group pre-post design in which there is no comparison group. This requires one to judge a program's effectiveness by seeing what changes occur in a treatment group over time. While this is easier to implement, it is weak from a statistical standpoint, since one cannot know which changes over time are due to the program and which are due to other, nonprogram factors.

While the comparison group design was judged in 1984, and again after a review in 1986, as on balance the best choice in this instance, it created certain problems. For example, many of the school site representatives would agree to provide data for a comparison group only if such students were anonymous. That is, we were allowed to gather data that were in school records but not to administer comparison group students either questionnaires or standardized tests. This is why the attitudinal data we have comes only from pre-post *program* student questionnaires.

Evaluation Problems

The collection of evaluation data about school-based programs is fraught with pitfalls and problems and requires more labor than often seems reasonable. In the worst case there are some programs that cannot be evaluated in a student outcomes sense, for a variety of reasons:

1. They have no precisely stated or storable objectives
2. Though they have reasonably precise objectives, these are not translatable into enumerable indicators of success
3. Though they have enumerable indicators of success, the data involved are impossible to collect due to expense, concerns of confidentiality, or bureaucratic barriers
4. The program or school staff are so defensive they refuse to cooperate with the data collection

Over the years, there were many examples of these problems in the SWAP and Academy evaluations. In most of the sites, there were fairly clear objectives, although translating these into enumerable indicators was difficult. As discussed earlier, program managers tended to state their objectives in terms of "helping at-risk youth," but they resisted measuring this through students' grades, credits, or test scores, feeling that their programs were not aimed primarily at academics. Attendance and retention in school were generally easier indicators about which to reach agreement; everyone agreed that these reflected program objectives. The advantage of all these indicators is that they are relatively inexpensive and unobtrusive forms of data to obtain because they already exist in school records.

Simple bureaucratic problems have been a common impediment in collecting these data. Most schools do not keep very good records. Often different systems exist for maintaining the same data between schools even within the same district. Attendance may

be kept by period or by day and is kept with varying degrees of accuracy by different teachers. Credits may be logged by a variety of unit systems. Grades may be filed by individual course or across all courses. All three may be kept by year or cumulative across years. Test scores may be recorded by "scale" score, grade equivalent, local percentile, national percentile, or normal curve equivalent. Retention data is particularly hard to obtain; schools know little of what happens to students once they leave, have varying and labyrinthine systems for categorizing such dropouts, and many even lack a clear definition of a dropout.

In addition to these problems of establishing clear data collection procedures for each indicator, there are invariably missing data for some students in all of the above categories. Most districts are in the process of computerizing their record keeping, with some data computerized and some not. In one district, I visited five offices, in several buildings, to obtain five categories of data (retention, attendance, credits, grades, and test scores), and when I mentioned this to a school principal his reaction was that I had wasted my time since the records at the high school were the only accurate ones anyway.

Defensiveness is another problem. No one likes to be judged. I have had to go to great lengths to explain the evaluation's methodology to program managers and to show them that it is fair and can be helpful to them in building program credibility. But ultimately, negative findings can destroy a program. Once one decides to have a serious evaluation, the program's survival can become tied to its results. It is not unusual for negative findings to result in challenges to the evaluation's methodology; I have never encountered such challenges when results are positive.

In one site, when the topic of evaluation was mentioned initially, the program director asked: "You're not going to use numbers on us are you?" I took this for defensiveness, but it turned out to relate to another problem. This manager had worked under CETA and was familiar with the ways in which performance tracking had led to the creaming of students to ensure the program's continuance and funding. He had understood that the Clark Foundation wanted to select the most at-risk students, and he felt that tracking student performance contradicted this goal.

This is invariably an issue, and the outcomes-oriented evaluation of the past three years has undoubtedly resulted in at least a slightly different approach to student selection at some sites, albeit one that few program managers would openly admit to. On the other hand, there is a positive side to the competitive pressure caused by an evaluation, as it forces the program to focus on student performance and to work hard at making a measurable difference in this regard.

There is also a relationship between the quality with which programs are run and how effectively they can be evaluated. There is always student attrition over time. If it is small, one can assume the program and comparison groups remain reasonably well-

matched. If it is large, as is typical of poorly run programs, statistical tests conducted two or three years later may be based on such small subsets of the two original groups as to have little meaning. This is true even in experimental designs. In short, poorly run programs with high dropout rates cannot be evaluated in as statistically precise a way as can well run ones.

Evaluation Guidelines

Underlying all this is a central question: has the Clark Foundation's investment in evaluation paid off? What was accomplished? The answer rests on a dilemma: evaluation is difficult, expensive, and imperfect. But despite all the problems, I believe it is essential—and so do policy makers and the public. Ultimately, the only way anyone can prove that a program really works is through evaluation. It is also central to the process of improving programs. To paraphrase Toynbee, those who fail to study their performances are condemned to repeat them.

In what specific ways has the evaluation served its purpose? There are several:

- It has strengthened the focus on accountability by the sites.
- It has provided proof, clearly in Chicago and to some degree in several other sites, that programs can succeed.
- It has strengthened the case for replications of programs and added to their credibility among policy makers.
- It has provided valuable feedback to sites in refining their programs and improving their performance.
- It has contributed to the evolution of clear program models.

In addition, the evaluations over the past several years have taught us a good deal about the evaluation process itself. These lessons pertain to evaluation designs, obtaining data from schools, the relationship between the evaluation and the programs, and what can ultimately be accomplished. Among these lessons are the following:

- The evaluation should fit the program. There is no point in conducting a student outcomes evaluation of a program that is brief and superficial in its treatment. There is no point in using academic indicators for a program that has no academic treatment.

- The evaluation should fit the stage at which the program exists. Process evaluation makes the most sense initially; a mixture of process and outcome is needed in the middle stages; outcome evaluation is probably of central interest eventually.
- Initially every program has start-up problems. Some form of systematic feedback to managers is needed to ensure the working out of initial problems and to give the program a fair trial.
- School-based data are hard to collect and invariably imperfect. This is not to say that they may not be better than other alternatives.
- Outcome evaluations are difficult and expensive. The more rigorous the approach undertaken, the more expensive the effort.
- Outcome evaluations affect programs. One can not conduct such an evaluation of a program without influencing it. Only undertake outcome evaluations if you are prepared to live by their findings.

CONCLUSION

In my view, the Clark Foundation is to be congratulated for its attempts in this realm. It is a tough field; there may be no tougher group to work with than disadvantaged urban youth. They are an embodiment of society's failures, and many of them carry deep resentments as a result. They are easy to ignore or abandon, as has largely been done at the federal level in recent years. While the foundation's efforts have not been an unalloyed success during the 1980s, there are many bright spots in its record. If these can be built upon and strengthened, it can rightfully boast a substantial legacy in the fight to bring fairness and equality to our society and its youth.

Appendix A

SWAP and Academy Program Treatments

The table on the next page uses five categories by which to classify program activities. These are:

- *Employability skills/job preparation*—Dress, speech, behavior, job search practice, etc.; those skills generally needed for any job
- *Vocational training*—Technical training related to a particular job field, such as electronics or computers, as provided in the Academies
- *Academic classes*—Basic skills classes (English, math, science, social studies) incorporated into the program
- *Enrichment activities*—Activities outside the classroom, such as tutoring, field trips, mentorships, counseling, and social and cultural events
- *Work experience*—Summer or school-year job at a company, provided by and related to the program

The table provides a summary of how the programs included in the evaluation during the 1985-86 school year stack up on these dimensions. The numbers are estimates of the total number of hours a given student spends in each of the five categories of activities in each program, across the years he or she spends in the program. By this time, Albuquerque and Philadelphia were not a part of the evaluation. Cleveland is omitted because there was no student treatment there yet. The Peninsula Academies is included, although it was evaluated elsewhere, since it was substantially supported by the Clark Foundation.

**Clark Foundation
SWAP and Academy
Program Treatments**
(hours devoted to various activities)

Site	Empl.Skills Job Prep.	Voc. Training	Acad. Classes	Enrich. Activ.	Work Exper.	Total
(treatment)						
Boston (1-2 years)	20	0	0	200	1300	1520
Chicago (3-4 years)						
Dunbar	160	1440	640	320	1700	4260
Farragut	410	204	720	250	480	2064
Denver (1 year)	400	40	360	140	1000	1940
East St. Louis (1-2 yrs.)	140	0	1080	300	600	2140
Oakland (1 yr.)	20	0	0	20	1050	1090
Palo Alto (3 yrs.)	220	480	900	450	720	2770
Pittsburgh (3 yrs.)	120	420	1080	240	720	2580
Portland (3 yrs.)	160	620	580	400	800	2560
Washington (3-4 yrs.)	220	990	2520	200	420	4350

Appendix B

Program/Comparison Group Differences, 1984-1987

In the table on the next page, I have attempted to provide a site-by-site summary of the student outcome results over the past three years. The table lists the indicators for which there have been statistically significant differences each year for the eight programs evaluated during that time. Five sites were not evaluated in terms of student outcomes: Albuquerque, Boston, and Oakland lacked academic components and were therefore not evaluable in a student outcomes sense; Akron never got off the ground; and Cleveland did not have a student treatment until the 1987-88 school year.

The evaluation has been widened and refined each of the past three years. In 1984-85, "attendance" and "GPA" were tracked. In 1985-86, "credits earned toward graduation" was added, and in 1986-87, "courses failed" was added. While standardized test scores have also been collected in the sites where they were available, they have shown a significant difference in only one school in Chicago, in math, and so are omitted.

The following qualifications are required to fully understand the table:

- The Peninsula Academies were evaluated elsewhere, and the data here vary in certain respects accordingly. For example, analyses were not performed separately by school but were by grade level. This program was not evaluated during the 1986-87 school year.
- Cooperation could not be obtained to select and track a comparison group in Philadelphia, and so no statistical tests were performed there.
- In Pittsburgh, no grade point average data were available; absence of differences on this variable reflects the unavailability of the data.
- In Washington, D.C., the program design at Dunbar High School precluded the possibility of identifying a matched comparison group, ruling out meaningful statistical tests at this school.
- "Negative" means that the comparison group outperformed the program group.

**Summary Table:
Statistically Significant Differences
Between Program and Comparison Groups
Across Three Years, 1984-87**

Site and School(s)	1984-85	1985-86	1986-87
CHICAGO			
Dunbar High School	Attendance, GPA	Attendance, Credits, GPA	Courses Failed, Credits, GPA
Farragut High School	Attendance, GPA	Attendance, Credits, GPA	Attendance, Courses Failed, Credits, GPA
DENVER			
North High School	Attendance- Negative, GPA	No Significant Difference	Courses Failed- Negative
West High School	Attendance- Negative	No Significant Difference	No Significant Difference
EAST ST. LOUIS			
East St. Louis H.S.	Attendance, GPA	No Significant Difference	Terminated
Lincoln High School	GPA-Negative	Attendance, Credits, GPA	Terminated
PENINSULA ACADEMIES			
	Attendance- Grade 12	Attendance- Grade 10 Credits-All grades	Not Evaluated
PHILADELPHIA			
	No Companion Group	Terminated	Terminated
PITTSBURGH			
	Attendance	No Significant Difference	Attendance
PORTLAND			
	Attendance, GPA	Credits	Courses Failed, GPA
WASHINGTON, D.C.			
Dunbar High School	Not Evaluated	Not Evaluated	Unclear
Woodson High School	Not Evaluated	Not Evaluated	Credits; Courses Failed- Negative

Appendix C ***Political Impact***

There are several criteria useful in determining the political success of these demonstration programs:

- The support given them by the school and district
- The support given them by the private sector
- Whether they continue to operate after foundation funding expires
- The attention they garner in the city and beyond
- Whether they are replicated within the city, state, or nation

In the table on the next page, I have made judgments about each site's success in this "political" sense, using a three-way distinction:

- "Clear" evidence of impact
- "Some" evidence of impact
- "No" evidence of impact

Summary Chart: Political Impact

Site	Clear	Some	None	Comments
Albuquerque		X		Citywide influence
Akron			X	Terminated early on
Boston	X			National model
Chicago		X		Growing influence in city
Cleveland		X		Some influence in district
Denver	X			11-site state replication
East St. Louis			X	Little influence
Oakland		X		Some influence in district
Palo Alto	X			20+ site state replication
Philadelphia			X	Little influence
Pittsburgh		X		Possible replication in city
Portland		X		Possible replication in city
Washington, D.C.	X			Wide influence in city