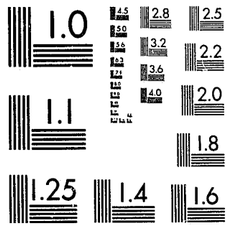


Center for Social Organization of Schools

Report No. 325
April 1982
THE SCHOOL ACTION EFFECTIVENESS STUDY:
FIRST INTERIM REPORT
Gary D. Gottfredson, Editor

ns Hopkins
versity

STAFF

Edward L. McDill, Co-Director
James M. McPartland, Co-Director

Karl L. Alexander
Henry J. Becker
Jomills H. Braddock, II
Shirley Brown
Ruth H. Carter
Michael Cook
Robert L. Crain
Doris R. Entwisle
Joyce L. Epstein
James Fennessey
Samuel A. Gordon
Denise C. Gottfredson
Gary D. Gottfredson
Linda S. Gottfredson
Edward J. Harsch
John H. Hollifield
Barbara J. Hucksoll
Lois G. Hybl
Richard Joffe
Debbie Kalmus
Helene M. Kapinos
Nancy L. Karweit
Hazel G. Kennedy
Marshall B. Leavey
Nancy A. Madden
Kirk Nabors
Deborah K. Ogawa
Donald C. Rickert, Jr.
Laura Hersh Salganik
Robert E. Slavin
Jane St. John
Valerie Sunderland
Gail E. Thomas
William T. Trent
James Trone

THE SCHOOL ACTION EFFECTIVENESS STUDY:

FIRST INTERIM REPORT

Gary D. Gottfredson, Editor

Report No. 325

April 1982

This report is sponsored in large part by Grant No. 80-JN-AX-0005 from the National Institute for Juvenile Justice and Delinquency Prevention, U.S. Department of Justice, and in part by Grant No. NIE-G-80-0113 from the National Institute of Education, U.S. Department of Education. The opinions expressed do not necessarily represent the position or policy of either agency. Published by the Center for Social Organization of Schools, supported in part as a research and development center by funds from the United Stated National Institute of Education, Department of Education.

Center for Social Organization of Schools
The Johns Hopkins University
3505 North Charles Street
Baltimore, MD 21218

Printed and assembled by the Centers for the Handicapped
Silver Spring, MD

Introductory Statement

The Center for Social Organization of Schools has two primary objectives: to develop a scientific knowledge of how schools affect their students, and to use this knowledge to develop better school practices and organization.

The Center works through five programs to achieve its objectives. The Studies in School Desegregation program applies the basic theories of social organization of schools to study the internal conditions of desegregated schools, the feasibility of alternative desegregation policies, and the interrelations of school desegregation with other equity issues such as housing and job desegregation. The School Organization program is currently concerned with authority-control structures, task structures, reward systems, and peer group processes in schools. It has produced a large-scale study of the effects of open schools, has developed Student Team Learning Instructional processes for teaching various subjects in elementary and secondary schools, and has produced a computerized system for school-wide attendance monitoring. The School Process and Career Development program is studying transitions from high school to post secondary institutions and the role of schooling in the development of career plans and the actualization of labor market outcomes. The Studies in Delinquency and School Environments program is examining the interaction of school environments, school experiences, and individual characteristics in relation to in-school and later-life delinquency.

The Center also supports a Fellowships in Education Research program that provides opportunities for talented young researchers to conduct and publish significant research, and to encourage the participation of women and minorities in research on education.

This report, prepared by the Studies in Delinquency and School Environments program, describes the interim results of the program's national evaluation of the Office of Juvenile Justice and Delinquency Prevention's (OJJDP's) Alternative Education Program.

School Action Effectiveness Study Staff

Center for Social Organization of Schools

Gary D. Gottfredson, Project Director
Denise C. Gottfredson, Associate Research Scientist
Michael S. Cook, Associate Research Scientist*
Deborah K. Ogawa, Research Analyst
Donald E. Rickert, Jr., Research Analyst
Helene Kapinos, Data Technician
Claire Skarda, Editorial Assistant
Stuart Gavurin, Research Assistant
Laura Gugerty, Research Assistant
Richard D. Joffe, Research Assistant
Robert Kirchner, Research Assistant
Hazel Kennedy, Secretary
Barbara Hucksoll, Secretary
Lois Hybl, Secretary

Social Action Research Center

J. Douglas Grant, Director of Formative Evaluation
Carol Yamasaki, Coordinator of Formative Evaluation**
Richard Carlton, Research Associate***
Deborah Daniels, Research Associate
Jane St. John, Research Associate

Federal Project Officers

National Institute for Juvenile Justice and Delinquency Prevention

Barbara Tatem, Project Officer

Special Emphasis Division, Office of Juvenile Justice and Delinquency Prevention

Monserrate Diaz, Project Officer
Cecilia Smith, Project Officer

---

* Joined the staff January, 1982.
** Left the staff January, 1982.
*** Left the staff September, 1981.

Staff

Technical Assistance and Training Division, Office of Juvenile Justice and Delinquency Prevention

Connie Walton, Project Officer

National Institute of Education

Bruce McKenzie-Haslam, Project Officer

Technical Assistance

Polaris Research and Development

Noel Day, Project Co-Director
David Sheppard, Project Co-Director
Dian Overbey, Project Associate

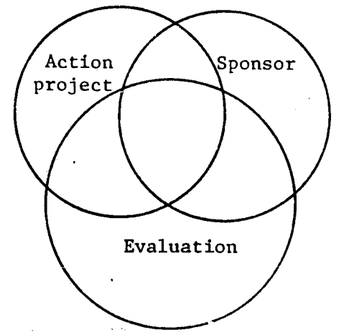
Preface

The School Action Effectiveness Study (SAES) is the national evaluation of the Office for Juvenile Justice and Delinquency Prevention's (OJJDP's) Alternative Education Program. The study is rooted in the perception that neither social action nor research will make progress without the collaborative effort of project implementers and researchers. Together, these two groups can create change and examine its consequences in settings where answers are needed and problems are real.

The study is also rooted in the notion that theory is an essential ingredient of both program development and evaluation research. Consequently, SAES aims to implement an action research model, in which project implementers work together with researchers, specifying theory-based research questions and designing their own evaluations as an aid to organizational self-study and continued project development.

Any large evaluation project creates tension. Multiple stakeholders, a collection of agencies and actors, and varied audiences with only partially overlapping--and sometimes divergent--interests present what will always be a challenge for evaluators. Because evaluation necessarily involves critical and skeptical examination, there may be no way to mesh these divergent interests completely. The approach taken in this project is to involve as many parties as possible in the evaluation enterprise. When the subjects of evaluation are collaborators in its development, the tensions may be reduced somewhat and evaluation feedback may be more helpful in project development.

We have been partially successful with this approach, but complete success has escaped our grasp. The reasons are many, and include limits on time, money, and our own talent and energy. Some of the more important reasons, however, are differences in the perspectives of the Federal sponsor, the various action projects, and ourselves. The Venn diagram shown below illustrates the problem. The various actors in this project have sometimes overlapping, and sometimes unique, goals or outlooks.



Preface

The primary sponsor (OJJDP) is first and foremost interested in delinquency prevention. It wants to evaluate demonstration projects--involving youth participation, organizational innovations, alternative reward structures, and individualized instruction (among other interventions)--creating institutional changes that may prevent delinquency. The action projects sometimes place less emphasis on delinquency prevention. They may be interested in educational achievement or truancy, or in continuing a program already in existence. Delinquency prevention is sometimes related to these interests even when it is not a primary aim. As evaluators, we are interested in assisting in the development of effective projects, critically assessing project effectiveness, and contributing to knowledge. Our critical perspective often results in approaches that diverge from the methods action projects find most comfortable. These projects often <u>assume</u> their interventions to be effective and prefer not to devote energy to the critical and sometimes painful scrutiny of those activities--especially in areas that are of interest to the sponsor and to science but that may be seen as tangential or even irrelevant by action project managers.

Our bias when encountering this tension has been to push as much as possible for a sound and thorough evaluation, in ways appropriate for each project. At the root of this bias is the conviction that the public deserves educational and delinquency prevention efforts whose effectiveness has been or can be demonstrated. This is especially true in a Federal demonstration program, in which the expenditure of public funds is justified by the evaluation of the resulting effort to learn how to develop and implement similar projects effectively. In short, we have little sympathy for the arguments that evaluation diverts effort from or detracts from programmatic work in these circumstances. Because the effectiveness of these action projects is unknown, the only ethical course is systematic evaluation. In addition, we assume that the evaluation of a project's efforts, the generation of knowledge about the consequences of a project's interventions, is a part of sound project administration and implementation.

We have not assumed that evaluating this program will be easy, and we are gratified that we have been as successful as we have been in translating our ambitions into reality. The excellent rapport and cooperation we have with the Federal agencies involved, and with most of the action projects, have been critical in this success.

This interim report summarizes some of what we have learned in the first year of the SAES. (For the most

part the report covers the period August 1980-August 1981.) Most action projects began implementing their interventions sometime during the past year, and all 17 projects are now out of their planning phases. Start-up activities are now behind us: Action projects have had up to a year to become accustomed to our expectations and methods, and we have had up to a year to acquaint ourselves with the action projects. We are pleased that evaluation is becoming routinized as an expected and well-understood part of the activities of most projects.

We are entering a second year of interaction with these 17 projects. In many cases, evaluation designs that are stronger than those possible in the first year are now being implemented. The next phase of this evaluation should be more productive in assessing project effects on students and schools.

Everyone is impatient for information about a project's "impact." This report is not the place to look for impact assessments. Here you will find information about a project's history, its start-up activities, and its successes and problems in implementation during its first year. You will find some organizational diagnoses, and some ideas about improving projects. Occasionally, you will find preliminary attempts to assess effectiveness. These preliminary attempts are not--nor are they intended to be--authoritative and conclusive statements. They are intended to provide information useful for project development.

Effective projects develop over time, incorporating feedback from their own observations and those of evaluators to become stronger. Provided that a project operates over a sufficient period of time, with a stable set of goals and guiding values, and with evidential pressure to guide the choice among alternative activities, an evolution that approaches an eventually stable "climax" program may be expected (Tharp & Gallimore, n.d.). This report is therefore directed primarily at project implementers and others, including OJJDP and its technical assistance contractor, who have a stake in fostering project development or in planning new initiatives. It is also directed, however, to the community of scientists engaged in the evaluation of social programs. The evaluation methods being developed in the School Action Effectiveness Study should be of interest to evaluators and students of evaluation.

Acknowledgments

This interim report is the product of many people's contributions, and it is possible to describe only some of these contributions here. Initial discussions leading to the proposal involved a small group: J. Douglas Grant suggested

using a program development approach and incorporating critical benchmarks as a component of the model. Denise Gottfredson conceptualized the information system, Lenore Campbell prepared a preliminary workplan on very short notice, and Gary Gottfredson provided the theoretical and methodological rationale for the project. This rationale was heavily influenced by the work of or discussions with LaMar Empey, Lee Sechrest, Joseph Weis, David Hawkins, and Travis Hirschi, among others. The general strategy owes much to the intellectual father of action research, Kurt Lewin, and is influenced by the example of Don M. Gottfredson in his work with criminal justice agencies. The evolution of the PDE model owes much to discussions with Carol Yamasaki, and the reactions of Social Action Research Center personnel in early staff training sessions, as well as to the the persuasive account of the ecology of program development and research provided by Roland Tharp and Ronald Gallimore.

This project is based in part on the contributions of thousands of students and teachers who shared their views about their schools and provided information about themselves. We hope they will be rewarded for their help by having their views heeded and acted upon.

Richard Carlton, Deborah Daniels, Denise C. Gottfredson, Jane St. John, Deborah K. Ogawa, Donald E. Rickert, Jr., and Carol Yamasaki worked long and hard with action project personnel in workshops, site visits, and on the phone to prepare for the surveys, to evolve Program Development Evaluation plans, and to draft project narratives. This report is possible because of their help and practical wisdom.

Ann Birdseye, Doris Coaxum, Barbara Dilligard, Hilda Gutierrez, Roland Patterson, Edward N. Whitney, Glen Bader, Hilda Irwin, Richard Smith, Charles Almo, Herman Steptoe, Thomas Corcoran, Mary Lewis, Preston Elrod, Vanita Vactor, Anita Batisti, David Bailey, Anadia Andrews, Tyrone Seals, Phyllis Betz, Ciorah Montes, Ivelissi M. Chardon-Zavala, Craig Overbeck, Phillip Cano, Cristobal Lopez, Roberto Duran, Nancy Cohen, Marilyn McKnight, Prentice Deadrick, Melvin Aaron, Dave Reiss, Lia Tepp, Joe Nathan, Mark Gilbert-Cougar, Nic Cooper, Sally Wisotzkey, Sonny Luster, William Kottman, Tom Leighty, Darlene Amyotte, Roy Mahoney, Pat Kenney, Marvin Dunn, Andrea Loring, Dolores Dunn, Michael Coleman, Pat Gahl, and Dorothy Martin were critical links between the real world and the evaluation, whithout whom there would be no School Action Effectiveness Study. They provided the theories that guided much of the instrument and scale construction, developed project plans using evaluation terminology, and made the action projects and data collection go.

Denise C. Gottfredson performed the superhuman task of coordinating and managing all of the data, Gary Gottfredson and Donald E. Rickert designed the survey instruments, and Deborah K. Ogawa and Donald E. Rickert put in many 18-hour days analyzing data under incredible time pressure. Computation at the Center for Social Organization of Schools is possible because of the work of Carol Weinreich.

Delbert Elliott, LaMar Empey, Joyce Epstein, J. Douglas Grant, Joan Grant, Travis Hirschi, and John Holland gave valuable advice on the development of the student questionnaire, not all of it taken. Raul Romero translated the student questionnaire into Spanish, and Dennis Dillon of Intran Corporation produced the optically scannable instruments.

We were ably assisted in managing and processing the data by Stuart Gavurin, Laura Gugerty, Richard D. Joffe, Helene Kapinos, and Robert Kirschner. Helene Kapinos coordinated the flow of an incredible amount of information. Lois Hybl helped keep the project orderly despite what often seemed to be incessant new demands; she maintained project files, typed manuscripts, and helped us to schedule our time.

Monserrate Diaz and Barbara Tatem of the Office for Juvenile Justice and Delinquency Prevention cleared the way for this project to proceed, and helped to resolve nearly countless problems along the way.

This report is edited by Gary D. Gottfredson, who was assisted by Claire Skarda. Ms. Skarda devoted many weeks to turning our nearly hopeless original manuscripts into grammatical and readable chapters. Gary D. Gottfredson wrote the chapters not attributed to others. Michael Cook, Deborah Daniels, Denise C. Gottfredson, Lee Sechrest, and Jane St. John made useful comments on a draft of several chapters. Opinions expresssed are the authors' or editor's, and do not necessarily reflect the position or policy of any agency or institution.

Organization of this Report

The remainder of this report is organized into two sections. The first of these discusses general topics that undergird or summarize the entire evaluation. Chapter 2 discusses the record of research in creating organizational change and delinquency prevention; it summarizes the weak nature of foregoing efforts, and argues that the defects of these earlier efforts must be overcome to increase the value of research and demonstration programs. Chapter 3 summarizes the conditions that lead to rigorous summative evaluation and argues that some common objections to creating these conditions can be overcome. Chapter 4 describes the

approach taken by the School Action Effectiveness Study, focusing on what we call the Program Development Evaluation Model. Chapter 5 discusses some of the measures used in SAES to provide the action projects with diagnostic information, and to measure delinquency and the important theoretical intervening variables. Chapter 6 provides a thumbnail sketch of the research designs for the 17 action projects, and summarizes the status of the evaluation and of the projects themselves. An overview of the results--all of which are formative rather than summative at the present stage of the evaluation--is also provided in Chapter 6.

The second section provides a narrative description of each action project. Most chapters were drafted by the field worker assigned to that project. Therefore, they generally have the benefit of having been given direct attention by the member of the evaluation team most familiar with the action project. At the same time, however, the involvement of multiple authors, each with a different background and perspective on evaluation, has led to some unevenness in presentation. Some authors have striven to include as much information as they could to foster project development and to characterize the projects thoroughly. Others have leaned strongly toward presenting the projects with which they work in a positive light, and have coped with the tension that could be created in the presentation of constructive criticism by downplaying that aspect of the report. The editorial process cannot eradicate the personal and stylistic differences that exist among the authors of these drafts. The reader is therefore urged to consider each of these narratives as a distinct essay, and to avoid making comparisons across projects on the basis of these individually drafted accounts. Many readers may be interested only in reading Part I. and then selectively dipping into chapters in the second section.

References

Tharp, R. G., & Gallimore, R. The ecology of program research and development: A model of evaluation succession. Unpublished manuscript. (Available from Roland Tharp, Department of Psychology, University of Hawaii, Honolulu 96822.)

Table of Contents

Part I

School Action Effectiveness Study Overview

Introduction to the School Action Effectiveness Study

The Alternative Education Program

The Office of Juvenile Justice and Delinquency Prevention (OJJDP) has funded 17 demonstration projects as part of a Program in Delinquency Prevention through Alternative Education. This OJJDP initiative is premised in part on the observation that individual delinquency is associated with a number of school-related or school-based problems, including disruptive classroom conduct, absenteeism, truancy, and dropout. An additional basis for fostering delinquency prevention through alternative education is found in a major theory of delinquency (Hirschi, 1969), in which commitment to educational or other conventional goals, attachments to teachers and the school, and belief in rules are viewed as bonds of social control which prevent delinquent behavior. Learning theory, especially social learning theory (Bandura, 1971), provides an explication of the ways in which these elements of the social bond may be strengthened by appropriate educational environments. Social learning theory also helps to explain how the influence of alternative school organization, and the influence of peers, teachers, and parents, can converge in preventing, or failing to prevent, delinquency. These theoretical perspectives find substantial support in the evidence provided by research; they concur in implying that alternative education programs can be structured in ways that will reduce delinquent behavior (Gottfredson, 1981b).

This outcome--primary, and to a certain extent secondary, prevention of delinquency--might be achieved in alternative education programs through their effects on the academic and social development of the youth involved.

The demonstration program is for the most part targeted at schools serving grades 6 through 12 in relatively high crime communities, with high rates of delinquency, dropout, suspensions, expulsions, absenteeism, and youth unemployment. Projects funded as part of this program were to be aimed at achieving (a) decreases in delinquent behavior in and around schools, (b) decreases in dropouts, suspensions, expulsions, and truancy, (c) increases in attendance, (d) increases in academic success in school with consequent increases in graduation rates, (e) improvements in the early post-schooling labor market experiences, or in the post-secondary training or education, of youth associated with participating schools.

The achievement of these objectives requires some reorganization of school

policies, practices, and environments. Specifically, the OJJDP program calls for achieving the following instrumental objectives to foster the attainment of the overarching program goals: (a) limiting or decreasing referrals to the juvenile justice system; (b) making school discipline fair and consistent while providing for due process; (c) increasing youth, parent, and community agency participation in school decision making to reduce student alienation and feelings of powerlessness; (d) decreasing the grouping of students according to inappropriate criteria (such as social class or race) which, accompanied by improved learning environments, should preclude labeling effects and stigmatization while enhancing educational success; and (e) providing a structure for learning that promotes educational and social development because it is tailored to realistic levels of performance for individual students.

A number of national advisory panels (President's Science Advisory Committee on Youth, 1973; Carnegie Council on Policy Studies in Higher Education, 1979; President's Commission on Law Enforcement and Administration of Justice, 1967) and delinquency researchers (Gold, 1978; Gottfredson, 1981a; Hawkins & Weis, 1980; Johnson, Bird, & Little, 1979; Hawkins & Wall, 1979) have argued that traditionally organized schooling, which focuses primarily on academic goals, does not provide opportunities to develop cognitive, interpersonal, and vocational competencies and may be inappropriate for many of today's students. Incorporating alternative educational options into school programs should provide more opportunity for development of such competencies or a better fit between student and school, thus promoting post-schooling vocational adjustment and preventing delinquency. Several authors (including Hawkins & Wall, 1979; Gottfredson, 1981a; McPartland & McDill, 1977; Johnson et al., 1979) have stressed the inability of conventional school reward structures to enable all students to experience success. This outcome--a failure of many students to be rewarded in school--decreases their stakes in conformity by decreasing their attachment to school and their commitment to educational goals. Youths who do not find school rewarding have little reason to conform. Consequently, alternative reward structures are one important feature of the alternative education provided in the action programs.

The OJJDP (1980) program announcement invited applications for action projects intended to alter school organization, climate, and educational practices. Specifically, the following characteristics were requested in the solicitation:

1. Projects are to provide opportunities for voluntary participation in alternative educational experiences aimed at "developing constructive interests relevant to (youths') environment" and promoting educational and psychosocial development. The focus of projects is to be on youth making transitions from elementary school to junior high and from junior to senior high school, although this focus need not be exclusive. Tracking, labeling, segregation, and stigmatization of students is to be avoided.

2. Projects are to specify goals and objectives aimed at achieving the OJJDP program goals and objectives discussed earlier.

3. Projects are to exemplify the following elements thought to be conducive to achieving these goals: (a) individualized instruction; (b) clear reward structures that supercede or supplement traditional classroom grades and that reward student improvement, incorporating a flexible array of rewards for different amounts of progress; (c) coalitions of school, community, business, parent, and youth leadership that attempt to change the educational environment; (d) comprehensive rather than piecemeal attempts to improve schools, involving a variety of relevant organizations and agencies; (e) use of peer and parental influence as a vehicle for accomplishing goals; (f) the training of school personnel and the creation and implementation of practices to increase positive interaction with and responsiveness to students; (g) small program size and favorable student-to-adult ratios; (h) strong, fair, consistent school governance and administration devoted to student growth; and (i) caring, competent teachers.

These OJJDP-generated project specifications constitute the first of three bases for an evaluation. The second basis is the theories of action which underly the project-specific goals, objectives, and intervention models each project develops. The third basis for the evaluation is the broader delinquency prevention and educational literature, which specifies some intermediary objectives that are important for delinquency prevention efforts.

Evaluation Aims

The overarching goal of the School Action Effectiveness Study is to create communicatable knowledge about delinquency prevention theory and practice. But a complex evaluation such as the School Action Effectiveness Study must accomplish many aims if it is to be effective. There are many audiences interested in delinquency prevention, but so far demonstration and evaluation efforts in the delinquency prevention area have amassed a dismal record.

The audience for the study includes project managers and their organizations, because they want (or sometimes need but do not want) feedback on their progress as one tool to use in developing their projects. The Federal sponsor is another audience with a direct and immediate interest in the evaluation because it has chosen alternative education as a promising area for research and development. The sponsor's mission is to contribute to knowledge in delinquency prevention and to develop prevention methods that can be suggested for broader implementation; it needs evaluation to accomplish this mission. The general public, keenly aware of what it perceives as widespread youth crime and disorderly schools, is a third audience, with both prurient and practical interest in a problem that affects everyday life. Evaluation researchers are struggling to develop paradigms for evaluation under difficult circumstances, and they are therefore eager to learn what others at the forefront of evaluation are doing. Researchers and theoreticians in sociology, psychology, and criminology have a direct and obvious interest in the knowledge gained through action research. Finally, Congress and, possibly, state and local legislative bodies want to know how effectively the public's funds are being spent, and what kinds of programs they should support in the future. Legislators and other policy makers are therefore important audiences.

As the chapter by Ogawa (this volume) makes clear, previous delinquency prevention efforts and their evaluations have been fraught with problems of incomplete implementation, weak evaluations, and lack of intermediary and outcome measures required to assess the efforts. Not only delinquency prevention programs suffer from these problems. Sarason (1971) describes the disappointing degree of implementation of attempted educational innovations such as the "new math." Whereas the developers of the innovation intended to alter the ways teachers interact with students, the major outcome was the use of some new math books. Lots of educational evaluations are, as Charters and Jones (1973) put it, evaluations of "non-events." The SAES must take steps to avoid evaluating non-events, and also to avoid the other problems from which earlier prevention evaluations have suffered.

The multiple consumers of this evaluation and the history of previous delinquency prevention efforts imply that SAES should accomplish the following goals:

1. The collection of sound measures of delinquent behavior, achievement, attendance, persistence in school, and vocational behavior, as well as measures of the relevant intervening theoretical variables

believed to be associated with these outcomes.

2. The establishment of evaluation designs that allow the most confident interpretation of results possible.

3. The documentation of project history, context, and conduct.

4. Documentation of the theoretical rationale underlying each project's interventions, and assessment of the plausibility of that rationale.

5. Documentation of project implementation in a way that allows assessment of the strength and integrity of that implementation, and replication of the interventions if warranted.

6. The development of sound project management plans to increase the effectiveness of each project and to help accomplish Goal 5.

7. The education of project managers in the utility of evaluation as a management tool and in the use of feedback about project implementation and effectiveness to foster project development.

8. The development of knowledge about conducting an evaluation.

9. The development of fundamental knowledge about delinquency prevention and education.

10. Explicit guidance for policy makers and future program implementers to increase the payoff from future expenditures of public funds.

These are the overarching goals of the School Action Effectiveness Study. These goals are being pursued in the context of a Federal research and development effort which provides further focus and structure for SAES. The remainder of this chapter explains some of the specific concerns of the evaluation and describes the interorganizational context of the evaluation.

SAES Concerns

Project Expectations about Evaluation

The OJJDP program announcement specified that action projects must explain how their approach would enable a national evaluation. It also required applicants to give assurances of their willingness to cooperate with a management information system. Furthermore, an appendix to the program announcement discussed the desirable features of an evaluation. The action projects selected, however, did not usually address these issues thoroughly. Some projects were surprised that they would be involved in serious summative evaluation, and would be asked to create arrangements to increase their projects' evaluatability. Other projects did not anticipate (despite the program announcement) that they

would be collaborating with a national evaluator. Considerable work was therefore required to (a) explain the elements of formative and summative evaluation to action project staffs, (b) obtain staff cooperation and gain access to the information required to conduct these evaluation activities, (c) demonstrate that evaluation can be helpful to project implementers, and (d) negotiate arrangements to increase project evaluatability. Unfortunately, the legacy of many previous so-called evaluations in the education and delinquency prevention fields is one of extensive miseducation about evaluation issues. Consequently, important short-term objectives for the SAES involved orientating action project staffs to a serious evaluation aimed not only at making hard-headed summative assessments of their projects, but also at assisting in project development.

Planning and Implementation

The history of evaluation research in delinquency prevention is replete with examples of programs in which the implementation was undocumented or not carried out as planned (Dixon & Wright, 1974; Krisberg, 1978; Ogawa, this volume). Knowing the fidelity with which program plans are implemented, the strength of the educational and social "treatment," and the context within which the program operates is essential for three reasons. First, any evaluation result--either positive or negative--is of little value unless the nature of the program is well described. Second, information derived from monitoring the activities and the implementation of plans is needed to strengthen the integrity of the program, and to detect unforseen consequences or potential breakdowns in project plans or the evaluation design. Third, negative results of summative evaluations have sometimes led observers to conclude that the interventions intended to be implemented do not work, whereas the interventions may not in fact have been implemented, implying a quite different conclusion. Knowledge of what was actually implemented is essential in drawing conclusions from tests of any planned intervention.

Project environment. A component of our work has been to describe the origins and development of the action projects. This includes a history of the practical and theoretical origins of the projects, accounts of the populations served, and description of the links between the schools or school systems and other agencies.

Strength and integrity of planned interventions. Assessment of the planning and implementation process consists of two distinct components (Sechrest, West, Phillips, Redner, & Yeaton, 1979). The first relates to considerations of the strength of the intervention plan. This is essentially a

matter of the construct validity of the measures intended to be taken in an intervention. In a medical analogy, if a person is suffering from a bacterial infection, treatment with a sufficient dose of an antibacterial agent may be deemed a construct-valid (and strong) treatment. Treating the same person with aspirin (in whatever dosage) would be deemed a weak treatment lacking in construct validity. No rules have been agreed upon for assessing the strength of programs such as the alternative education action projects. Several procedures are available, however. These include (a) analysis of the plausibility of the plans' theoretical premises, and determination of how closely the specifics of the plans are linked to delinquency prevention theories; (b) expert judgments about the likelihood that the project as specified will produce the desired outcomes; and (c) comparisons of the intended programs with the range of current or past efforts at delinquency prevention (in this way a program that was otherwise unremarkable but resembled a previous ineffective effort might be judged a weak program). In addition to a theoretical basis, parameters involved in making assessments of strength include staff stability or qualifications, intensity and duration of treatment, focus of effort, clarity of plans, and the extent to which the plans involve different responses to different persons (e.g., individualized instruction). In general, replications of previously tested or well engineered interventions, comprehensive attempts to cope with the multiple causes of a problem, treatments with clearly spelled out treatment protocols or implementation manuals, or primary prevention efforts that affect a substantial proportion of an environment's inhabitants are likely to be judged stronger than those that lack these characteristics.

The second aspect of assessing program implementation relates to the integrity or fidelity with which plans are implemented. Clear plans are more likely to be implemented with fidelity than diffuse plans, fuzzy promises, or vague project descriptions. Some components of implementation that must be monitored or observed are (a) staffing patterns (including experience, training, numbers, and stability), (b) methods used to select, admit, or reject the youth involved in each project and each of its components, (c) the differential assignment of youth to alternative programs, or the basis for individualization of instruction, (d) the nature, duration, circumstances, and frequency of services to individuals or groups, (e) methods used to determine who (including students) is involved in implementation, (f) the interventions' elements and their duration, (g) the

degree of project staff commitment, (h) project supervisory and management practices, and (i) curricular materials, individualized education plans, lesson plans, diagnostic protocols, treatment plans, and the like.

The importance of this aspect of assessing implementation can scarcely be overestimated. The scope of the alternative education action projects, encompassing as they do many distinct components, makes the faithful implementation of all plans unlikely. A failure to obtain sound evidence about the strength and integrity of these prevention projects could lead to erroneous conclusions about the efficacy of the delinquency prevention ideas behind these projects. It could be a mistake, for example, to conclude that reorganized educational reward structures do not help in preventing delinquency (or in promoting career development) on the basis of negative summative evaluation results. Specifically, this conclusion could be a mistake if there were no solid evidence that reward structures were actually altered in systematic ways. Equally important, even if a summative evaluation implied that a project had been effective, in the absence of sound information about what actually was done the project would provide little basis for its replication at a new site. Such a project would provide only the shakiest guide to others who wish to implement a similar program.

Changes in Policies, Practices, Procedures, and their Implications

The OJJDP Alternative Education Program hopes to alter school policies and practices dealing with dropouts, school disruption, and delinquency, and to determine the implications of those changes for the school and its students. Empirical, theoretical, and practical considerations (Gottfredson & Daiger, 1979; National Institute of Education, 1978; Toby, 1980; Howard, 1978) implicate poor or inconsistent school disciplinary practices in the failure to prevent disruption in schools. Furthermore, evidence implies that youths who will drop out of school are more often disciplinary problems and experience more absolute or relative academic failure while still in school (Elliott & Voss, 1974; Hirschi, 1969; Gottfredson, 1981a; Hawkins & Weis, 1980; Johnson, 1977; Gold, 1978). School practices, policies, and procedures for coping with or responding to disruptive behavior--especially the fairness, firmness, and consistency of with which rules are applied--are of great importance in preventing delinquent behavior and other forms of misconduct. Improvements in this area may be expected to pay off in terms of reduced delinquency. Similarly, altering schools' responses (McPartland & McDill, 1977) to youth who have difficulty in

coping with traditional academic programs (by providing individualized curricula, by rewarding and encouraging the development of a wider variety of social skills and vocationally related competencies, and providing a more extensive range of rewards and responses) may be expected to increase learning, promote psychosocial development, and decrease delinquency and dropout. Thus, improved school governance, altered curricula, and increased responsibility may all directly or indirectly lead to decreased student misconduct and increased school retention rates. These changes, if they occur, must be documented by the SAES.

Youth and Parent Participation

The evaluation aims to determine the effect of action programs on youth and parent participation in school activities. An increase in such participation is expected to prevent delinquent behavior.

In Hirschi's (1969) theoretical account, youth involvement in conventional activities and commitment to conventional goals or pursuits are important bonds to society which serve to control behavior. And, youth involvement in school activities carries with it the opportunity for increased interaction with peers and teachers, an outcome that also may serve to increase stakes in conformity. Hirschi (1969) marshalls some empirical support of this theoretical perspective.

Parental involvement in school activities may also have salutory effects. Recent reviews of the use of home-based reinforcers as an aid to the classroom management of disruptive behavior (Barth, 1979; Atkeson & Forehand, 1979) imply that cooperation of parents in providing backup reinforcers is useful. Gaining that cooperation is a major practical problem. In addition, ample testimonial evidence (McPartland & McDill, 1977; Hawkins & Wall, 1979, p. 25) implies that parent involvement may be important.

The efficacy of increased parental or student involvement in school decision making is more dubious. Gottfredson and Daiger (1979), in a reanalysis of the Safe School Study data, conclude that no evidence that such participation is related to school disruption exists in that study of over 600 schools. They accord with the original National Institute of Education (1978) and Hawkins and Wall (1979) assessments in this regard. Despite considerable testimonial evidence that such participation may be important (summarized in Hawkins & Wall, 1979), little firm evidence or carefully articulated theory implies that student or parental participation in decision making is a promising strategy to reduce delinquency. Available evidence is based, however, on the analysis of natural variation. Because

schools typically do not involve students in major ways in decision making, this evidence does not show the consequences of major youth involvement in decision making.

The notion that student participation in conventional activities such as athletics, band, clubs, student government, and the like may prevent delinquency has more support in the literature. Here, theory (Hirschi, 1969) and research (Gottfredson & Daiger, 1979) converge in implying that such participation may be important. Indeed, typical explanations of the often-observed association between school size and delinquency (Hawkins & Weis, 1980) involve arguments about the lowered opportunity for participation or involvement in large schools (Garbarino, 1978; McPartland & McDill, 1977; Wiatrowski, Gottfredson, & Swatko, 1980).

Youth and parent participation is an intermediate outcome. As such, it could be considered a measure of the strength and fidelity of an intervention. The overall goal of reduced delinquency is expected to come about as a result of increased participation. Thus, there are two important evaluation questions: (a) To what degree are projects characterized by student and parent participation? (b) Does participation appear to contribute to delinquency reduction, and to decreases in truancy and absenteeism? An ancillary set of questions related to the second of these is whether increased participation produces the theoretically expected increase in attachment, commitment, and belief.

School Achievement: Social, Academic, and Vocational Skills; and Postsecondary Vocational Behavior

One set of objectives for the evaluation is to determine the effects of the alternative education programs on (a) educational performance, (b) social, academic, and vocational development, and (c) the transition between secondary education and work, postsecondary education, or vocational training.

Experimental evidence and theory predict that altered reward structures will influence educational outcomes. Specifically, interventions involving the reorganization of academic rewards, so that all students are rewarded in proportion to their educational improvement rather than in accordance with their performance relative to other students, hold much promise. Slavin (1980) recently reviewed the literature showing that cooperative team learning is a powerful way of harnessing peer group interaction to promote learning at the same time that it improves students' satisfaction with the educational process and increases learning according to standardized achievement tests. Various strategies are described in the

experimental literature Slavin discusses, but they have two themes in common. First, groups of students--teams or classrooms--are rewarded in some way on the basis of group performance. Second, each student can contribute to the performance of the group regardless of his or her current level of academic performance. This is accomplished either by assigning points to the team based on improvements in individual performance above each student's baseline performance, or by structuring competition so that students of approximately equal ability compete with each other. Points are then credited to the team based on this structured competition. Because students compete with others of approximately equal ability, all students contribute points to the team in approximately equal proportion. This is in sharp contrast to the traditional classroom system in which some students never are rewarded or perceive themselves as contributing to the performance of a valued peer group or class.

Such learning structures have never been evaluated for their effects on delinquent or disruptive behavior. But theory predicts that such programs, which resemble what the OJJDP has called for in its Alternative Education Program, will reduce delinquency if implemented with sufficient strength and fidelity. The existing evaluations of these programs show (a) increased academic performance, (b) increased self-esteem, (c) increased peer friendships, and (d) increased satisfaction with school. When translated into Gold's (1978) and Hirschi's (1969) theoretical terms, these outcomes imply increased self-esteem (leading to decreased need to use delinquent behavior as an ego-defense), and increased attachment to schools and to peers. Johnson et al. (1979) summarize additional evidence that the kinds of peer group and reward structures created by such interventions may be effective.

Some action projects are attempting to "individualize" instruction by using differential educational treatments, such as alternative curricula or teaching styles, or by developing learning plans based on an individual diagnosis. A recent comprehensive review by Cronbach and Snow (1977), which analyzes work seeking to establish and use knowledge about interactions between student characteristics and instructional treatments, confirms the utility of this approach, although progress in this area is not as great as would be hoped. Attempts at "individualizing" instruction, based on research by Fizzell (1979) or by Hunt (1974), are described in the OJJDP solicitation's background paper by Hawkins and Wall (1979); these attempts are examples of strategies that programs could attempt to implement and which show at least some promise.

The evaluation task is threefold: to describe and document the implementation of the interventions used by the action projects, to assess the contribution of these activities to student academic performance, and to assess the contribution of improved academic performance or skills to delinquency reduction.

Vocational and Educational Development

These and other aspects of the alternative education projects may influence social and vocational skills. Altered student participation rates in a variety of school activities, more experience with a broader range of curriculum (some of which is directed to vocational and interpersonal development), and the altered nature of peer group interaction and reward structure may all contribute in some degree to these outcomes. Krumboltz (1978), for example, has spelled out a social learning theory of vocational development which implies that such influences should alter individual competencies and inclinations to pursue various careers. The development of vocational and interpersonal skills should increase youths' stakes in conformity, and thus prevent delinquency.

The task of assessing the effects of these projects on the transition from secondary school to post-secondary employment, training, or education is a difficult one. Longitudinal studies extending beyond the anticipated three- or four-year duration of the evaluation would be helpful here. The reason for this is that only relatively few students will have experienced a project for three years and accumulated any post-secondary work, training, or educational experience in this time span. For students experiencing fewer than three years of an alternative education program, the intervention will probably lack sufficient strength to produce substantial effects. Despite these limitations, evaluation tasks include documenting, insofar as is possible, the educational and vocational plans of students leaving secondary school, and assessing the contribution of various program components on those early career outcomes.

Dropouts, Suspensions, Expulsions, Truancy, and Delinquency

Assessment of the effects of the alternative education projects on rates of dropout, suspensions, expulsions, truancy, and delinquency is a major goal of the evaluation. Improved educational experiences as a result of the alternative education programs, if implemented with sufficient strength and integrity, should influence these outcomes.

Academic performance is a strong correlate of delinquent behavior in and out of school (Bachman, O'Malley, & Johnston, 1978; Elliott & Voss, 1974; Empey & Lubeck,

1971; Hawkins & Weis, 1980; Gottfredson, 1981a). Truancy and dropping out of school also appear to form part of a constellation of behavior of which delinquency is a frequent concommitant. Interventions that prevent delinquency may also be expected to influence these outcomes, both on the basis of empirical evidence and on the basis of theory (Hirschi, 1969), which postulates that attachment to school is an important ingredient in delinquency prevention.

Program Models Most Effective for Different Types of Youth Under Different Conditions

The evaluation also seeks to determine which types of alternative education models appear most effective for different types of youths, and under what conditions. This is a challenging task. Clear and confident answers to this set of evaluation questions will almost certainly not be forthcoming. We are limited to information derived from 17 projects, which are attempting to implement different interventions, with differing degrees of fidelity, and which are serving populations that differ. Most of these projects have not been set up to permit the unambiguous search for the kind of statistical interactions demanded by these evaluation questions. And the history of the search for statistical interactions in quasi-experimental research is a discouraging one (D. Gottfredson, 1981).

Despite these difficulties, it is undoubtedly worthwhile to dredge the evidence from the 17 action projects for clues about what works best, for whom, under what conditions. Certain interventions are most effective for certain types of individuals. For example, a youth who performs poorly in school is expected to receive few rewards from traditional education and thus to have low attachment to school and little commitment to traditional educational goals. Empirical evidence supports this generalization (Hirschi, 1969; Sewell, Haller, & Portes, 1969; Bachman et al., 1978). Other students, because they receive rewards and perform well in traditional schooling, are already attached to school and committed to educational goals. A program designed to alter reward structures may be effective in decreasing delinquency, truancy, and dropout among the former group but may perhaps have no effect on the latter group.

This is but one example of the theory-derived expectations that can guide a search for interaction effects. Other theories appear to predict that enhancing self-esteem through alternative education may be more important for youths with little social control than for those with high levels of social control (cf. Gold, 1978). A thorough theoretical approach (cf. Glaser, 1977) to the search for interventions most effective

for particular subgroups appears to be the most fruitful way of pursuing this evaluation goal.

The "under what conditions" part of this evaluation question hinges on issues of implementation. Projects with the most plausibility (or in the terms used earlier, the most strength, and those which are implemented with most fidelity), will likely create the "conditions" that are most effective. Other conditions that merit scrutiny have been mentioned earlier. They include staffing patterns and stability, resources, extensiveness and duration of services, community characteristics, and the interorganizational environment within which the project operates. Learning about the conditions necessary to create effects is likely to involve a boot-strap operation, in which clues to the conditions necessary come from the theory-based examination of the nature of effective interventions.

Evaluation, the Sponsor, and the Action Projects

The Alternative Education Program is sponsored by the Office of Juvenile Justice and Delinquency Prevention, with supplemental funding provided through OJJDP by the Department of Labor. Three divisions of OJJDP are involved directly in this program. First, the Special Emphasis Division has programmatic responsibility for the grant awards made to the 17 action projects listed in Tables 1 and 2. Second, the Technical Assistance and Training Division has responsibility for providing assistance in project development, and works through contractors to do so. Initially, the Westinghouse National Issues Center was assigned these technical assistance tasks as part of its larger contract to provide assistance for OJJDP's Delinquency Prevention Research and Development efforts. In recent months, Westinghouse was replaced by Polaris Research and Development in this role. Third, the National Institute for Juvenile Justice and Delinquency Prevention is responsible for the evaluation. The Institute made a grant to the Johns Hopkins University to perform this evaluation, and the University subcontracted part of the work to its collaborator, the Social Action Research Center. In short, a total of 23 organizational entities are directly involved in this effort. The participation of each is essential to the successful conduct of the evaluation.

The degree of collaboration and cooperation among these groups has been exemplary. A major difficulty facing many evaluations is rivalry, or a lack of coordination among the various agencies involved. In this evaluation, however, the staffs of OJJDP's Institute, Special Emphasis Division, and Technical Assistance and Training Division have met frequently with us and with Polaris to coordinate

activities, assist in each other's efforts, share information, discuss problems, and plan solutions. This coordination has led to some blurring of the action projects' perceptions of the roles of the evaluation and the agency, a confusion that has on occasion created small problems. The most salient of these are (a) action projects sometimes using the evaluation staff as a conduit for issues more properly addressed to their Federal project officers, and (b) action projects sometimes assuming that a Federal agency concern is an evaluation priority as well. Occasionally, this collaboration has also resulted in some resentment when a project officer emphasized the importance of cooperating with the evaluation, or when evaluation problems or information was shared with a project officer.

These minor problems are outweighed by the positive contributions of this joint approach. The national evaluation would not have been possible without a coordinated approach involving all three divisions of OJJDP. Because persons submitting proposals for action projects under the alternative education initiative did not really expect to be evaluated rigorously, and because we had to discuss touchy issues (such as collecting data about the criminal behavior of students, implementing evaluation designs, and monitoring project implementation activities) with action agencies, the evaluation would have been torpedoed from the beginning had we not had the backing and understanding of the Special Emphasis Division, which is responsible for monitoring the action projects.

There are well-known hazards in collaborating with a Federal sponsor on the evaluation of a program in which it has a vested interest. One hazard is noted by Cronbach and associates (1980, p. 4), who say, "Insofar as information is a source of power, evaluations carried out to inform a policy maker have a disenfranchising effect." This may occur when "only the officials know what is going on." We hope to minimize the danger in this area in the present evaluation. By design, this evaluation is intended to foster the development of more effective projects by directly involving action project managers in conducting the evaluation and by feeding information back to those managers as a project development tool. Furthermore, the open dissemination of evaluation reports is a mechanism for informing all audiences of findings that may be a source of power. A second hazard is discussed by Gottfredson (1978) and by Weiss (1975). Because evaluation takes place in a political context in which multiple stakeholders are competing for the allocation of resources, there is the possibility that evaluation may be misused in policy debates. Although this danger may never be completely avoided, we aim to minimize it by open communication and due scientific circumspection. We are above all scientists, albeit scientists tinkering in the area of social action and social policy. Therefore we seek to guard against overly effusive statements that are not based on sound evidence. This may disappoint both Federal sponsors and action agencies, but it is the only defensible course.

Hazards also confront an evaluator collaborating with project implementers. The first is akin to the hazards of collaboration with a Federal sponsor: Every action project wants an evaluation to make it look good, and wants to use evaluation results in its political struggle for survival. And no action project wants an evaluation to be used--as is so often the case--to justify its demise. Therefore project implementers are at once eager for and afraid of evaluation. The second hazard is that, through sympathetic interaction with persons earnestly trying to do good, the evaluator may contribute to the misuse of evaluation. Our approach to both of these hazards is to acknowledge that they are threats, and again to seek umbrage in scientific skepticism and open discourse.

A third hazard may exist. Commentators have divergent views about the proper role of an evaluator in influencing project process. One view holds that it is not appropriate to intervene in the conduct of a demonstration project because such evaluator intervention would probably not be available in more wide-scale subsequent adoptions of a program model, thus threatening the external validity (generalizability) of the evaluation. Another view holds that formative evaluation is an essential aspect of the evaluator's role, and that evaluators should intervene by providing information according to the action research model. Perloff (1979), who discusses the divergent views using the OE/NIE/AIR/ERS experience with the "Cities in Schools" program as an illustration, leans toward the first view. We endorse the second, especially in the present case. The OJJDP program is a research and <u>development</u> project. Development of models is clearly an appropriate goal, given the state of delinquency prevention theory and practice at present. The nation's experience with delinquency prevention attempts is so fraught with weak programs (Ogawa, this volume) that excessive worry about evaluator intervention leading to inappropriate transportability conclusions are premature. The primary tasks at present are to demonstrate that <u>some</u> interventions can work and to learn how to facilitate the implementation of such interventions.

References

Atkeson, B. M., & Forehand, R. Home-based reinforcement programs designed to modify classroom behavior: A review and methodological evaluation. Psychological Bulletin, 1979, 86, 1298-1308.

Bachman, J. G., O'Malley, P. M., & Johnston, J. Adolescence to adulthood: Change and stability in the lives of young men. Ann Arbor, Mich.: Institute for Survey Research, 1978.

Bandura, A. Social learning theory. Morristown, N. J.: General Learning Press, 1971.

Barth, R. Home-based reinforcement of school behavior: A review and analysis. Review of Educational Research, 1979, 49, 436-458.

Carnegie Council on Policy Studies in Higher Education. Giving youth a better chance: Options for education, work and service. San Francisco: Jossey-Bass, 1979.

Charters, W. W., Jr., & Jones, J. E. On the risk of appraising non-events in program evaluation. Educational Researcher, 1973, 2, 5-7.

Cronbach, L. J., & associates. Toward reform of program evaluation. San Francisco: Jossey-Bass, 1980.

Cronbach, L. J., & Snow, R. E. Aptitudes and instructional methods: A handbook for research on interactions. New York: Irvington, 1977.

Dixon, M. C., & Wright, W. E. Juvenile delinquency prevention programs: An evaluation of policy related research on the effectiveness of prevention programs. Nashville, Tenn.: Office of Educational Services, Box 60, Peabody College for Teachers, 1974.

Elliott, D. S., & Voss, H. L. Delinquency and dropout. Lexington, Mass.: Lexington, 1974.

Empey, L. T., & Lubeck, S. G. Explaining delinquency. Lexington, Mass.: Heath, 1971.

Fizzell, R. L. The truants alternative program: An evaluation report for the State Board of Education, Illinois Office of Education. Macomb, Ill.: Western Illinois University, 1979.

Garbarino, J. The human ecology of school crime: A case for small schools. In E. Wenk & N. Harlow (Eds.), School crime and disruption: Prevention models. Washington, D.C.: National Institute of Education, 1978.

Glaser, D. Concern with theory in correctional evaluation research. Crime and Delinquency, 1977, 23, 173-179.

Gold, M. Scholastic experiences, self-esteem, and delinquent behavior: A theory for alternative schools. Crime and Delinquency, 1978, 24, 290-308.

Gottfredson, D. C. Black-white differences in educational attainment. American Sociological Review, 1981, 46, 542-557.

Gottfredson, G. D. Practical and ethical concerns in collaborative research with criminal justice decision makers. Paper presented at the annual meeting of the American Psychological Association, Toronto, 1978.

Gottfredson, G. D. Schooling and delinquency. In S. E. Martin, L. B. Sechrest, & R. Redner (Eds.), New directions in the rehabilitation of criminal offenders. Washington, D.C.: National Academy Press, 1981. (a)

Gottfredson, G. D. Schooling and delinquency prevention: Some practical ideas for educators, parents, program developers, and researchers (Report No. 304). Baltimore: Johns Hopkins University, Center for Social Organization of Schools, 1981. (b)

Gottfredson, G. D., & Daiger, D. C. Disruption in six-hundred schools: The social ecology of school victimization (Report No. 281). Baltimore: Johns Hopkins University, Center for Social Organization of Schools, 1979.

Hawkins, J. D., & Wall, J. S. Alternative education: Exploring the delinquency prevention potential. Seattle: University of Washington, Center for Law and Justice, 1979.

Hawkins, J. D., & Weis, J. G. The social development model: An integrated approach to delinquency prevention. Seattle: University of Washington, Center for Law and Justice, 1980.

Hirschi, T. Causes of delinquency. Berkeley: University of California Press, 1969.

Howard, E. R. School discipline desk book. West Nyack, N.Y.: Parker, 1978.

Hunt, D. E. Conceptual level matching model. In D. E. Hunt & E. V. Sullivan (Eds.), Between psychology and education. Hinsdale: Dryden, 1974.

Johnson, G., Bird, T., & Little, J. W. Delinquency prevention: Theories and strategies. Washington, D.C.: U.S. Department of Justice, Law Enforcement Assistance Administration, Office of Juvenile Justice and Delinquency Prevention, 1979.

Johnson, R. E. Delinquent behavior: The development and text of a causal model (Doctoral dissertation, University of Washington, 1976). Dissertation Abstracts International, 1977, 37, 4640-A. (University Microfilms No. 77-590)

Krisberg, B. Preliminary report of the national evaluation of prevention. San Francisco: National Council on Crime and Delinquency, 1978.

Krumboltz, J. D. A social learning theory of career selection. In J. M. Whiteley & A. Resnikoff (Eds.), Career counseling. Monterey, Calif.: Brooks/Cole, 1978.

McPartland, J. M., & McDill, E. L. (Eds.) Violence in schools. Lexington, Mass.: Lexington, 1977.

National Institute of Education. Violent schools--Safe schools: The safe school study report to Congress. Washington, D.C.: Author, 1978.

Ogawa, D. K. Delinquency prevention: The record of accomplishment. This volume.

Office of Juvenile Justice and Delinquency Prevention. Program announcement: Prevention of delinquency through alternative education. Washington, D.C.: Author, 1980.

Perloff, R. (Ed.) Evaluator intervention: Pros and cons. Beverly Hills, Calif.: Sage, 1979.

President's Commission on Law Enforcement and Administration of Justice. Task force report: Juvenile delinquency and youth crime. Washington, D.C.: U.S. Government Printing Office, 1967.

President's Science Advisory Committee on Youth. Youth: Transition to adulthood. Washington, D.C.: U.S. Government Printing Office, 1973.

Sarason, S. B. The culture of the school and the problem of change. Boston: Allyn and Bacon, 1971.

Sechrest, L., West, S. G., Phillips, M. A., Redner, R., & Yeaton, W. Introduction. In L. Sechrest, S. G. West, M. A. Phillips, R. Redner, & W. Yeaton (Eds.), Evaluation studies review annual (Vol. 4). Beverly Hills, Calif.: Sage, 1979.

Sewell, W., Haller, A., & Portes, A. The educational and early occupational attainment process. American Sociological Review, 1969, 34, 82-92.

Slavin, R. E. Cooperative learning in teams: State of the art. Educational Psychologist, 1980, 15, 93-111.

Toby, J. Crime in American public schools. Public Interest, 1980, 58, 18-42.

Weiss, C. H. Evaluation research in the political context. In E. L. Struening & M. Guttentag (Eds.), Handbook of evaluation research (Vol. 1). Beverly Hills, Calif.: Sage, 1975.

Wiatrowski, M. D., Gottfredson, G. D., & Swatko, M. K. Classifying school environments to understand school disruption (Report No. 295). Baltimore: Johns Hopkins University, Center for Social Organization of Schools, 1980.

Table 1

Action Project Names, Locations, and Award Amounts

Organization Name	Location	Number	Amount	Awarded	Beginning	Ending
Prevention of Delinquency through Alternative Education	St. Croix, Virgin Islands	80JSAX0030	$267,812	8/13/80	9/1/80	8/31/82
Individualized Integrated Alternative Education (Otro Camino)	Playa Ponce, Puerto Rico	80JSAX0031	$692,609	8/14/80	9/1/80	8/31/82
Project PREP	South Bronx, New York	80JSAX0032	$1,196,967	8/14/80	9/1/80	8/31/82
Compton Action Center--Youth Development Alternative School	Compton, California	80JSAX0033	$607,682	8/15/80	9/1/80	8/31/82
Peer Culture Development	Chicago, Illinois	80JSAX0034	$606,194	8/15/80	9/1/80	8/31/82
Jazzmobile Alternative Education Arts Program	Harlem, New York	80JSAX0035	$668,019	8/15/80	9/1/80	8/31/82
George I. Sanchez Alternative Education Program	Houston, Texas	80JSAX0036	$529,583	8/15/80	9/1/80	8/31/82
Prevention of Delinquency through Alternative Education	Kalamazoo, Michigan	80JSAX0037	$268,315	8/15/80	9/1/80	8/31/82
Positive Action through Holistic Education	Charleston, South Carolina	80JSAX0038	$883,508	8/15/80	9/1/80	8/31/82
Alternative Education Program	Hayward, Wisconsin	80JSAX0039	$539,779	9/10/80	9/1/80	8/31/82
Project RETAIN	Chicago, Illinois	80JSAX0043	$1,088,983	9/10/80	9/1/80	8/31/82
Student Training Alternatives through Urban Strategies	Pasadena, California	80JSAX0044	$594,902	9/10/80	9/1/80	8/31/82
Prevention of Delinquency through Alternative Education	Sewell, New Jersey	81JSAX0012	$602,601	12/16/80	12/15/80	12/14/82
Milwaukee Youth Employment Center	Milwaukee, Wisconson	81JSAX0014	$1,156,105	12/16/80	12/1/80	11/30/82
Plymouth-Canton Alternative Programs	Plymouth-Canton, Michigan	81JSAX0013	$462,779	12/16/80	12/1/80	11/30/82
Academy for Community Education	Miami, Florida	81JSAX0015	$477,940	12/16/80	12/1/80	11/30/82
Project Together	St. Paul, Michigan	81JSAX0022	$300,765	1/19/81	1/1/81	8/31/82

Table 2

Organizational Characteristics of Alternative Education Action Projects

Organization Name	Location	Type	Primary Mission
Virgin Islands Department of Education, Elena Christian Junior High School	St. Croix, Virgin Islands	Public school system	Education
Dispensario San Antonio, Inc.	Playa Ponce, Puerto Rico	Not-for-profit service	Social service and community development
Community School District 9	South Bronx, New York	Public school system	Education
Joint Center for Community Studies, Compton Action Center for Youth Development	Compton, California	Not-for-profit service organization	Community development and social service
Peer Culture Development, Inc.	Chicago, Illinois	Not-for-profit service organization	Youth development
Jazzmobile, Inc.	Harlem, New York	Not-for-profit service organization	Arts education
Association for the Advancement of Mexican Americans	Houston, Texas	Not-for-profit service organization	Social service
Western Michigan University, Department of Sociology	Kalamazoo, Michigan	State university	Post-secondary education and research
Charleston County School District	Charleston, South Carolina	Public school system	Education
Lac Courte Oreilles Tribe	Heyward, Wisconsin	Tribal council	Tribal government
Chicago Board of Education	Chicago, Illinois	Public school system	Education
Constitutional Rights Foundation	Pasadena, California	Not-for-profit service organization	Law-related education
Educational Improvement Center-South	Sewell, New Jersey	State government sanctioned educational service	Technical assistance
Jewish Vocational Services	Milwaukee, Wisconsin	Not-for-profit service organization	Vocational rehabilitation
Plymouth-Canton Community Schools	Plymouth-Canton, Michigan	Public school system	Education
Institute for Innovative Interventions	Miami, Florida	Not-for-profit service organization	Human service delivery
St. Paul Public Schools	St. Paul, Minnesota	Public school system	Education

Preventing Delinquency: The Record of Accomplishment

Deborah K. Ogawa

Many delinquency prevention programs have been initiated, but few have been carefully evaluated. It is not known whether most of these programs have positive, null, or negative effects. Among the few programs that have been evaluated reasonably carefully, there have been some positive results. Examples include Alexander and Parsons' (1973) short-term behavioral intervention with delinquents and their families, Reid and Patterson's (1976) attempts to reduce aggression and stealing behaviors by modifying the reinforcement pattern within the family setting, and Barth's (1979) review of 24 studies utilizing home-based reinforcement to alter behaviors in the school setting. The Alexander and Parsons study and the Reid and Patterson work adhered to rigorous evaluation designs by using a randomized control group and a matched control group, respectively. They are also exemplary because of the clear plausibility of the intervention models.

The Record

These examplary projects are rare, indeed. Dixon and Wright (1975) reviewed 95 delinquency prevention reports published after 1965 and concluded that there is a paucity of evidence about the effectivenesss of existing programs, and that when evaluations have been conducted, few projects have shown positive significant results. They attributed part of the problem to unclear project goals and objectives, and to difficulties encountered in implementing rigorous designs and in obtaining meaningful measurements in a fluid action program setting. Hawkins and Wall (1980), in describing an alternative education program to reduce delinquency in Florida, also identified design, measures, and data collection and analysis procedures as three major problems in evaluation research.

Krisberg (1979) reviewed 16 exploratory delinquency prevention programs funded by the Office for Juvenile Justice and Delinquency Prevention. After one year of operation, only one project had been able to implement even a quasi-experimental design. Most of the projects could not be evaluated in terms of their effectiveness because of problems in data collection and lack of comparable control groups. In addition, none of the 16 projects had articulated a useful theory about delinquency in their catchment areas or had delineated the ways in which their services would reduce the problem. Goals were often too ambi-

tious or too ambiguous, and were not clearly related to the problems the projects were to address. Krisberg also concluded that the failure to achieve these goals was due in part to incomplete planning.

The story is the same everywhere one looks. Wall, Hawkins, Lishner, and Fraser (1980) reviewed 36 "model" juvenile delinquency prevention programs, only two of which utilized a rigorous evaluation design. Although many of the programs suggested positive effects, the designs were not rigorous enough to exclude other rival hypotheses about the reasons for the results. Janvier, Guthmann, and Catalano (1980) rated 52 evaluations of drug abuse prevention programs for youth on the basis of their methodological rigor. In fewer than half (46%) of the programs would the evaluation designs allow conclusions to be drawn. In addition, only half of the evaluations used at least one outcome measure related to drug abuse. Of the 52 projects, only 9 had an adequate design and at least one outcome measure related to drug use.

Terpstra (in press) reviewed 52 articles published between 1965-1980 on the evaluation organizational development efforts that involved the collection and analysis of quantitative data. Reports were rated on six dimensions:
(a) sampling strategy,
(b) sample size, (c) control group utilization, (d) use of random assignment,
(e) measurement strategy, and (f) significance level. Results show that 5% of the studies indicated a negative significant effect; 23%, a mixed or nonsignificant effect; and 67%, a positive significant change. Although over half of the articles indicated a significant positive effect, there exists an inverse relationship between the degree of methodological rigor of the evaluation and the degree of successful outcome; i.e., studies purportedly showing positive effects were generally lower in methodological rigor. These deficiencies in methodological rigor make the conclusions about the utility or effectiveness of the interventions questionable. The Terpstra analyses are valuable because they illustrate the potential for misguided enthusiasm about a project's perceived effectiveness in the absence of careful study.

In summary, many previous evaluations of delinquency prevention and organizational change programs do not yield dependable conclusions about the programs' effectiveness. Design flaws are one factor limiting the dependability of a study's conclusion. The use of irrelevant measures poses a second problem in delinquency prevention evaluations. Some evaluations do not include any delinquency measures at all. In addition, measures that are used are often poorly operationally defined, and there is a dependence on only one

source of data, usually official records. More multiple-measure evaluations--evaluations that include self-reported delinquency measures to supplement official records--are needed to reduce the ambiguity of evaluation results (cf. Hawkins & Wall, 1980).

Theory in Evaluation

Another major problem with delinquency prevention evaluations has been that many of the previously evaluated programs have not implemented truly plausible interventions based on a theory (cf. Glaser, 1980). Programs often fail to articulate a theory of delinquency prevention that would provide a conceptual framework for project planning, implementation, and evaluation. Consequently, evaluators have to ferret out, post hoc, underlying theoretical assumptions. Programs that do not utilize theory add little to the development of knowledge in the area of delinquency prevention, and implementation often suffers because project implementers have no standard against which to assess their interventions.

Due to the weaknesses and limitations of past evaluation research in the area of delinquency, innovative approaches are needed. One promising approach is the action research model (Lewin, 1947). Action research is the study of actions as a method for advancing both knowledge and practice, through a cycle of problem analysis, planning, execution, evaluation, and replanning (Sanford, 1970). The first step in the action research model is planning, which involves defining the problem and then examining ways to resolve it in relation to the available resources. Once an overall plan has been formulated, the next stage, execution, begins. A fact-finding step then evaluates the action that has been executed. This evaluation provides an opportunity to gain new insights about the plan's strengths and weaknesses and serves as a basis for the next step, modification of the plan. Action research thus involves a spiral of steps: It allows continuous improvement of a program through evaluation of the results of each action to provide a rational basis for planning the next action. In order to gain insight into a process, change must be created, followed by observation of the new effects and dynamics. As a result of this cycle of activities, programs should become more effective.

Although this process was widely used in the 1940's by the Research Center for Group Dynamics and the Commission on Community Interactions, action research has never been influential in psychology or the social sciences in general (Sanford, 1970). Very recently there has been a small resurgence in the use of the action research model (Lenrow, 1970; Grant & Grant, 1970; Hoff, 1970). Scriven (1967), in distinguishing between formative and

summative evaluations, began to approximate the Lewinian model of action research. Recent writing on the evaluation of delinquency prevention efforts is now moving in this direction. Hawkins and Wall (1980) delineate standards for evaluating delinquency prevention programs that include not only a summative, or outcome, component, but a formative, or process monitoring, component as well. Summative evaluation involves rigorous research designs, standardized measures, and an appropriate research time frame which provides for longitudinal follow-ups. Process monitoring describes the program, making replication possible. This description includes the context of the program and the selection of participants, and it includes the documentation of intervention strategies.

Evaluator-Implementer Collaboration

Krisberg (1979), Dixon and Wright (1975), Janvier et al. (1980), and Glaser (1980) all advocate the use of theory in guiding program development and evaluation. The greater the degree of specificity of the theory, the more readily identifiable are a program's set of measurable goals. These goals then should provide the framework around which intervention strategies are to be tailored. In addition, project implementers should be involved as collaborators in research. Again, in accordance with the Lewinian action research model, involvement is an important aspect of group decision-making since it minimizes resistance to considering the problems and possibilities of an objective and it allows expression of several alternatives. Thus, involvement of program implementers in the research process may reduce the program staff's resistance to, and anxiety about, being evaluated. More important, the underlying assumption is that project staff members are more effective change agents if they participate in and have a stake in the decision-making and research process.

Good Examples

Many of these recommendations have already been incorporated in evaluation studies. Empey and Lubek (1971) and Empey and Erikson (1972) have integrated sociological theory in delinquency prevention intervention efforts. Their work included a formulation of the theory, intervention strategies, and methods for empirically testing these formulations. Alexander and Parsons (1973), recognizing the paucity of demonstrable intervention effects in the psychotherapy literature, incorporated a strategy in their evaluation of family therapies which involved four main goals: (a) presentation of a clear description of intervention techniques, (b) process evaluation, i.e., describing and evaluating the behavioral changes in family process expected from the intervention, (c) summative

evaluation, i.e., using clearly defined and nonreactive behavioral criteria to evaluate the effectiveness of the intervention, and (d) incorporation of controls for maturation and professional attention (p. 219). This study was able to effectively utilize a stringent experimental design with three groups: families receiving treatment, families receiving alternative forms of family therapy, and families receiving no professional treatment. This study implemented a strong design and demonstrated positive treatment effects in the reduction of recidivism in deliquent teenagers.

The School Action Effectiveness Study is an attempt to use the experience of previous programs and their evaluations to anticipate and avoid as many pitfalls as possible. It aims to assist in clarifying goals and theory, and their linkages with short-term or intermediary objectives and the interventions aimed at bringing these objectives about. And it also aims to provide workable structures for managing project implementation and evaluation according to the action research model. It combines formative evaluation or project development with rigorous evaluation.

References

Alexander, J. F., & Parsons, B. V. Short-term behavioral intervention with delinquent families: Impact on process and recidivism. Journal of Abnormal Psychology, 1973, 81, 219-225.

Barth, R. Home-based reinforcement of school behavior: A review and analysis, Review of Educational Research, 1979, 49, 436-458.

Dixon, M. C., & Wright, W. E. Juvenile delinquency prevention programs: An evaluation of policy related research on the effectiveness of prevention programs. Washington, D.C.: National Science Foundation, 1975.

Empey, L. T., & Erikson, M. C. The Provo experiment: Evaluating community control of delinquency. Lexington: Lexington Books, 1972.

Empey, L. T., & Lubeck, S. C. The Silverlake experiment: Testing delinquency theory and community intervention. Chicago: Aldine-Altherton Press, 1971.

Glaser, D. The Interplay of theory, issues, policy, and data. In M. Klein & K. Teilman (Eds.), Handbook of criminal justice evaluation. Beverly Hills: Sage, 1980.

Grant, J., & Grant, J. D. Client participation and community change. In D. Adelson and B. Kalis (Eds.), Community psychology and mental health. Scranton: Chandler, 1970.

Hawkins, J. D., & Wall, J. S. Alternative education: Exploring the delinquency prevention potential. (Report of the National Juvenile Justice Assessment Center). Washington, D.C.: U.S. Government Printing Office, 1980.

Hoff, W. New health careers demonstration project. Berkeley, Calif.: Institute for Health Research, 1970.

Janvier, R. L., Guthmann, D. R., & Catalano, R. F., Jr. An assessment of evaluations of drug abuse prevention programs. (Report of the National Juvenile Justice Assessment Center). Washington, D.C.: U.S. Government Printing Office, 1980.

Krisberg, B. Executive summary of the national evaluation of prevention. Preliminary report. San Francisco: National Council on Crime and Delinquency, 1979.

Lenrow, P. Strengthening early education: Collaboration in problem solving. Berkeley: Berkeley Unified School District, 1970.

Lewin, K. Group decision and social change. In T. M. Newcomb & E. L. Hartley (Eds.), Readings in social psychology. New York: Holt, Rinehart, & Winston, 1947.

Lundman, R. J., & Scarpitti, F. Deliquency prevention: Recommendations for future projects. Crime and Delinquency, 1978, 24, 207-220.

Reid, J. G., & Patterson, G. R. The modification of aggression and stealing behavior of boys in the home setting. In A. Bandura & E. Ribes (Eds.), Behavior modification: Experimental analysis of aggression and delinquency. New Jersey: Lawrence Erlbaum Associates, 1976.

Sanford, N. What happened to action research? Journal of Social Issues, 1970, 26, 3-23.

Scriven, M. The methodology of evaluation. In Perspective on curriculum evaluation (American Educational Research Association monograph series). Chicago: Rand McNally, 1967.

Terpstra, D. Relationship between methodological rigor and reported outcomes in OD evaluation research. Journal of Applied Psychology, in press.

Wall, J. S., Hawkins, J. D., Lishner, D., & Fraser, M. Juvenile Delinquency prevention: A compendium of 36 program models. (National Center for the Assessment of Delinquent Behavior and its Prevention). Seattle: Center for Law and Justice, 1980.

Making Inferences about Project Effectiveness

Once a project has implemented some plausible intervention intended to influence student attitudes, behavior, or development, assessing the consequences of that intervention becomes important. Making this assessment is not always easy. Young people are growing and changing all the time. Rates of participation in delinquent behavior apparently rise and then fall with age. Scholastic competencies usually grow over time, but at different rates for different people. Students make new friends and abandon old ones, and every parent knows that his or her child's tendency to conform or rebel is different at different stages of development. Isolating the influence of some specific experience, intervention, or set of interventions is therefore difficult.

Making inferences about the causes of some difference in student outcomes--about the effects of planned interventions--is, however, a major goal of evaluation. Put another way, an aim of a thorough evaluation is to determine whether an observed difference in student behavior or attitudes (if any difference is observed at all) can

---

I am grateful for comments by Deborah Daniels, Denise C. Gottfredson, and Jane St. John on a draft of this chapter.

reasonably be attributed to a specified intervention. Certain conditions make the search for the effects of an intervention easier; other conditions preclude making any confident inferences. Those conditions are the topic of this chapter. An excellent discussion related to this topic exists elsewhere (Cook & Campbell, 1979), and readers may wish to see that source for an elaboration of some of the points made here.

Rival Explanations

When an educational or other intervention has been executed with fidelity, the evaluation task focuses on learning the consequences of that intervention. In practice, of course, no project can wait until after the intervention has been implemented to begin work on this task. Conditions must be established at the outset to allow the conclusion that observed outcomes were brought about by by the intervention, rather than by something else happening at the same time, or by the natural course of development. Project implementers and evaluators ignore the establishment of these conditions at their peril; causal interpretations of observed outcomes are dubious unless rival explanations can be ruled out.

Suppose, for example, that an alternative education project involving group

and individual counseling were to be implemented. The counselors implementing the treatment believe that only students willing to participate fully and amend their previous conduct are amenable to this treatment.<1> Therefore, only students who express an earnest willingness to commit themselves to the project become involved in the counseling activities. Under these circumstances, counselors often make claims for the effectiveness of their intervention by comparing the past behavior of these students with their behavior during or shortly after counseling. Or, they may claim effectiveness based on a comparison of students receiving treatment with apparently similar students who did not become involved in treatment. These claims are on shaky ground. Any differences may be due to the desire of the individuals involved to reform, or to maturation, and may have nothing to do with the treatment. The rival explanations are as good as the one the counselors wish to make.

Consider a second example. Educators are conducting a project involving individualized education plans developed by a specialist in collaboration with their students' regular classroom teachers. The basic idea is to make a diagnosis of each student's needs and specify achievable academic and behavioral objectives; the specialist is to serve as a kind of ombudsman to promote the educational welfare of the students. Classroom teachers are asked to refer to the project students for whom these special services seem appropriate, and they are given a list of criteria to guide them in making referrals. Referral criteria call for students who exhibit mild behavior problems, such as difficulty in impulse control or persistent truancy, or for students whose classwork performance is at a level below the teachers' expectations. The educators administer tests (and collect certain other information) prior to, during, and after students' involvement in the individualized education. Because scores on these tests increase, the educators claim that the treatment is effective. But these claims are questionable. Scores on educational tests almost always go up over time, especially when any instruction is occurring. Therefore, gains cannot necessarily be attributed to the treatment.

Suppose, then, that a novice educational researcher wanted to examine this same project more closely. The novice tries to construct a "control" group using students of the same race, sex, grade, and age as the students involved in the individual education treatment, drawn from the same classrooms as the students referred to the project, i.e., by "matching." Fortunately, the same tests administered to the project clients were also administered to students not involved in the project. Behold, the students receiving individualized attention show smaller gains than the "control" students. The novice concludes that the treatment was actually harmful (students would have learned more if not involved in the project). Perhaps not. Recall that the students referred to the treatment were performing below expectations, were exhibiting behavior problems, and were often truant. These students may be expected to show educational growth that was slower than the growth of other "matched" students in the same class in the absence of any special intervention. Thus, difference in expected educational growth rates is an explanation with as much credibility as that of the novice educational researcher.

Sophisticated measurement and statistical techniques can sometimes help sort out the evidence about an intervention's effects under difficult conditions. But these non-experimental efforts to make inferences are plagued with uncertainty. Few such efforts that capture the attention of other methodologists go unchallenged for long. In short, positive steps to assure an intervention's evaluatability are essential if confident statements are to be made about the project's effectiveness in terms of its intended outcomes.

Randomization and Alternatives

A number of methods allow reasonably confident inferences. Of these the true experiment and some quasi-experimental methods such as single-subject or ABA designs, and regression discontinuity designs, require some degree of experimental control over the timing of treatment. Other quasi-experimental methods, such as interrupted time-series designs, require a large number of observations over a long period of time, together with clear-cut changes in some environmental influence at a particular point in time. (These methods and others are described in Cook & Campbell, 1979.) It is always wise to consider whether any of the various quasi-experimental designs are likely to be credible in ruling out rival explanations of a particular outcome.<2> True experiments are usually administratively simpler, fraught with fewer technical difficulties, and more generally understood.

True experiments usually involve the random assignment of individuals, classrooms, schools, time periods, or other units to alternative treatments. No single procedure is always effective in guarding against all rival interpretations, but randomization is a helpful general purpose mechanism.<3> When two or more groups are created through randomization, they are equivalent within the limits of random sampling

error; methods for estimating the size of this error are available. Had true experiments been conducted in the hypothetical cases described earlier, confident conclusions would have been possible--provided that certain other conditions necessary for inference were also present.

Other Conditions Making for Confidence

To make confident interpretations of evaluation results, three important additional conditions are required: adequate statistical power, sound measurement of the outcomes of interest, and complete information.

Statements about treatment effects made by evaluators and other scientists are probabilistic. Generally, the degree of confidence one may have in a conclusion is expressed by indicating the probability of this outcome occurring by chance if the intervention were completely ineffective. This is what statisticians mean by "significance." A significant result is one that is unlikely to occur by chance. Many scientists, as well as lay persons, are confused by the distinction between the size of a difference and its statistical significance. Large differences in the average delinquency rates or educational achievement test scores between two groups can be nonsignificant. And, small differences can be significant. In most delinquency prevention interventions, as indeed in most educational interventions, treatment effects are likely to be small. Detecting such small differences with confidence--and understanding the paradox of nonsignificant large differences--requires a consideration of statistical power.

Power is the probability of detecting a difference of any given size that hypothetically exists. The most important principle involved is that the probability of detecting a true difference with conventional significance tests increases as the number of experimental units (students, classes, time periods, or schools) increases. If the true difference is large, a smaller number of experimental units is required to detect it with a given level of probability. Thus, when small treatment effects are likely, large numbers of people must be given the treatment to make a significant result probable. When effect sizes are large, smaller samples can be used and still provide a reasonable probability of detecting the effects. In most delinquency prevention evaluations, large sample sizes are needed to gain statistical power because intervention effects are likely to be small.

The second additional condition necessary for making confident statements about an intervention's effects is the sound measurement of the outcomes that

may be influenced by the intervention. The review by Ogawa (this volume), as well as our experience in trying to obtain sound measures of academic achievement, delinquent behavior, and other potential outcomes in this evaluation, implies that this condition can often be difficult to meet. There is no way to confidently conclude that an intervention prevents delinquency when no good measure of delinquent behavior is available. Measurement issues are discussed more thoroughly in another chapter.

The third necessary condition, complete information, is also often difficult to meet because school populations are transient, and because students in high risk of delinquent involvement are often truant or drop out of school altogether. In addition, some students for a variety of reasons never receive the intended treatment in full form. The necessity of obtaining outcome measures for these individuals is frequently overlooked. Attrition weakens an evaluation by effectively diluting the treatment. And, if information for some individuals is not available, a number of equally plausible rival explanations for outcomes may exist, thwarting confident interpretation.

Some Common Objections

Educational practitioners, counselors, and social service workers often object to establishing the conditions necessary for making confident inferences about the services they provide or about their organizational change efforts. These objections take many forms:

1. "I know this intervention to be effective, and therefore evaluation is unnecessary."

2. "Randomization is unethical."

3. "Asking students to report about their behavior or school is demeaning."

4. "Evaluation is too much work; it detracts from other programmatic efforts, or makes the intervention difficult to implement as intended."

5. "Evaluation threatens the stability of the project by creating problems for its manager or powerful others in the manager's environment."

6. "Evaluation restricts the project's freedom of action in developmental stages."

Evaluation Is Unnecessary

"I know this treatment works. Evaluation is not necessary."

No intervention involved in the Alternative Education Program has been demon-

strated to be effective in preventing delinquency. Most have not been demonstrated to be effective in achieving any of the other program goals. Some interventions being contemplated, e.g., the Student-Team-Learning approach being considered in Charleston, have reasonably been shown effective in increasing student satisfaction, learning, and positive peer relations in implementations conducted by skilled researchers. Replications under more typical conditions are lacking. Others, such as the FOCUS approach, have been subjected only to unreplicated examination involving a limited range of potential outcomes. Still others, such as PLATO and a host of other interventions, have never to the best of our knowledge been satisfactorily evaluated at all. (We acknowledge that our standards for a satisfactory evaluation are considerably higher than the standards of those who market or otherwise disseminate these products.) Even were it true that some evaluation had found an intervention to be effective in preventing delinquency, the replicability of that intervention and its results would be an important evaluation question in a demonstration program. Evaluation is therefore necessary.

Randomization Is Unethical

"Refusing or delaying this service to allow for its evaluation is unethical."

The denial of some service, known to be effective, for evaluation purposes when the aggregate harm to the individual outweighs the resulting aggregate benefit is unethical. This is a fundamental tenet of utilitarian ethics. Rule ethics arguably imply further that denial of a service, known to be beneficial to an individual, may be unethical regardless of the aggregate harm or benefit. But the denial of a service when its efficacy is unknown is not unethical. Indeed, when effects are unknown, the ethicality of administering the treatment is doubtful, especially if the treatment is not under serious evaluation.

When the effectiveness of a treatment is unknown, the alternative to experimenting with people is fooling around with people. Consider again some examples. Peer or group counseling is a plausible intervention because it recognizes the powerful influence of peers on a student's behavior. Some reluctance to assign students randomly to this kind of counseling has been encountered. This resistance is based in part on concerns about denying a needed service to individuals who would be randomly assigned to a waiting list or control group. The excellent background paper for the Alternative Education Program Announcement (Office of Juvenile Justice and Delinquency Prevention, 1980, Appendix 3) makes clear, however, that the appropriate approach to peer

counseling interventions is an experimental one. "Given the growing popularity of peer counseling and the likelihood that some alternative programs will use it, it is essential to rigorously assess its effects in alternative education programs. It cannot be assumed that positive results will be found" (p. 24, emphasis added).

The ethical approach to such interventions is to evaluate them. As the American Psychological Association's Task Force on Evaluation and Accountability (1978) put it, "In the vast majority of cases the only really ethical position lies in providing the public with effective services or services whose effectiveness is under systematic evaluation" (p. 305).

In a second example, alternative schools may seek to keep dropouts or potential dropouts in school. Again, resistance to randomization has been based in part on concerns about denying a needed service to individuals who would randomly be assigned to a control group. Yet the benefits of continued schooling for youths with high dropout potential have not been demonstrated. Delinquent behavior typically falls off when youths drop out of school and may even be lower during summer recess from school (Gottfredson, 1981). Some delinquency researchers have commented that "dropout is a satisfactory solution" for some youths (Elliott, 1966), and that we should rethink the appropriateness of trying to keep youths in school as long as possible (Glaser, 1975, p. 47). Bachman, Green, and Wirtanen (1971) characterize dropout as a symptom of more basic organizational problems, rather than as a problem for the individuals who drop out themselves. Treating this symptom rather than the underlying problems "may in this instance do more harm than good for two reasons. First, the treatment has some unpleasant side effects . . . . Second, treating the symptom may distract us from the more basic problems" (p. 179).

The importance of learning the effects--which may be positive, negative, or mixed--of keeping potential dropouts in school cannot be overestimated. This is an issue with tremendous policy relevance. The ethical route lies in the rigorous evaluation of any program designed to keep potential dropouts in school.

The case for randomization when the effectiveness of an intervention is unknown has been elaborated elsewhere (Boruch, 1975; Gottfredson, 1978). Randomization is often considered a model of fairness in allocating benefits or risks (Feinberg, 1971) and has much to recommend it on that basis alone when the consequences of a treatment are unknown. Surely the once commonly accepted and widely practiced medical treatment involving blood letting

would have been abandoned sooner, saving countless lives, had anyone performed the necessary evaluation (cf. Eisenberg, 1977).

Asking Certain Questions Is Demeaning

"Why should innocent youngsters be asked if they have committed crimes? Why should they be asked if school stinks? Isn't this demeaning?"

Questions have to be asked of students to learn how interventions affect them. Student self-reports on their conduct are one, albeit imperfect, method of learning about their delinquent behavior. Self-report measures are well studied (Hindelang, Hirschi, & Weis, 1981), and we know they have useful degrees of validity for an evaluation. Students' perceptions of their schools and their experiences in schools are important in assessing school climate for diagnostic purposes, for characterizing the school environment, for assessing project effectiveness. And because attitudes are important variables that theory implies mediate between plausible interventions and delinquency or dropout, measures of these attitudes are also important.

Asking these questions is apparently not demeaning. The overwhelming majority of students (81.5%) who completed the School Action Effectiveness Student Questionnaire reported that it was very or somewhat interesting. Discussions with students imply that they generally appreciate being asked their opinions. Ignoring the messages students give us when they answer those questions may be demeaning; asking the questions is not.

Evaluation Gets in the Way

"We have not budgeted for a staffer to collect those data. Spending effort on developing the evaluation takes precious time away from getting our work done. We cannot meet our quota for persons served if we have to establish a control group too."

These are real problems. More than a decade ago a prison warden (Park, 1965) told a story that is modified slightly and repeated below:<4>

Once upon a time there was an alternative education project director who was riding a tiger. By holding on with both hands and struggling very hard, he could steer the tiger just the slightest bit. The project director's best efforts were not enough to prevent the tiger from taking an occasional nip out of his leg.

One day an evaluation researcher visited and said, "I see you are riding a tiger," thus demonstrating her keen insight into the project director's plight. At

that moment, the tiger took another bite from the director's leg.

The researcher observed solemnly, and presently issued a finding: "You know, that tiger is biting you, and seems to enjoy it. Someday he will eat you all gone--unless, that is, you avail yourself of my services."

"You know about riding tigers?"

"No," said the researcher, "but I have extensively studied Siamese kittens, and I am sure the principles are the same."

Although he was able to devote only a small part of his attention to the researcher's statement, being almost fully absorbed in efforts to control the tiger, the project director made an executive decision. Since things were going badly at the moment, he thought he had little to lose by getting the researcher's help. "Fine," he said, "come aboard and give me a hand."

"Not so fast," answered the researcher, "we scientists can't go slapdash into things. In the first place, we must develop an evaluation design, administer some questionnaires, and develop a data retrieval mechanism. And in the second place, we must find a quieter tiger. Simply impossible to properly study such a rambunctious creature."

"But this is the only tiger I have, and if I let it go he will run around eating a lot of innocent students, and scare them into dropping out of school."

"Bosh," replied the researcher. "You are being rigid. If you will loosen your grip a little, we can randomly assign it to individualized training. We'll provide it with caring and concerned trainers. If we find that it eats fewer students, or even if an intervening variable is influenced, we will have a result. Of course if it eats more students, we will still have a result--opposite direction naturally. So we can't lose, can we?

"Well," said the project director, who was getting confused by the researcher's logic, "I'm not sure about that. Help me get this tiger under control, then we can think about evaluation."

"Now, now," answered the researcher, "you can't just start these things without preliminary study. We must devise a management information system, get some data from the tiger's point of view,

and prepare some feedback. Right now, we cannot even describe what you two are doing."

"Chomp," said the tiger.

All right, all right," said the project director, "describe already--but hurry."

"Now, let's see--I would say at the 2% level of confidence that we could have some preliminary results that would point up areas requiring intensive study in about five years, plus or minus 2.3 years." The researcher looked pleased.

"I don't like to be an obstructionist," replied the administrator, "but I have this tiger to cope with now, and I am not sure I will be around in 5 plus or minus 2.3 years."

"You must realize," said the researcher, "that we must develop criteria, select subjects, and make sure the results are not due to the use of catnip rather than the tiger's individualized training plan. Besides, we will generate valuable information to help the next project director. You wouldn't want the next poor soul who comes along to have the same problems you are having with this tiger, would you?"

"Hmmmmmmmm," hummmmmmed the project director, "that doesn't sound at all practical. However, while you are working on the evaluation design, perhaps you can help me shift my grip a little. Down a ways and a bit to the right should do it."

"I'm pretty committed," said the researcher, "to developing a rigorous national evaluation, but I can give you a few minutes of consultation. Here, hold these data collection forms and I will help you optimize your grip."

The project director reached out for the research tools, momentarily loosening his grip. The tiger promptly turned and ate him all gone.

The researcher regarded the scene sadly. "Just when he was coming around to my frame of reference."

The project director, may his soul rest in peace, had a point--several points. But his problems both antedated and went beyond the problems created by the researcher. This administrator was not in control. An organization must be stable and have enough control of its course to be able to make sound decisions and investigate the consequences of its action, even though

the crisis may stimulate him to do <u>something</u>.

An organization must be willing to dedicate a share of its attention and resources to inquiry if it is to engage in and use research--indeed, if it is to develop at all. A failure to budget adequate resources for evaluation is a direct indication that an organization does not value the contribution that evaluation can make to project development. It is a problem that must be overcome; otherwise, the attempt to evaluate must be abandoned. Evaluation research does require a commitment of resources.

In the disorganized case of the project director riding a tiger, devoting attention to evaluation is indeed hazardous. Few project directors would argue, however, that the work they wish to accomplish is riding a tiger. If the project director is out for a tiger ride, research will get in the way. But as Lewin (1946) put it:

> In a field that lacks objective standards of achievement, no learning can take place. If we cannot judge whether an action has led forward or backward, if we have no criteria for evaluating the relation between effort and achievement, there is nothing to prevent us from making the wrong conclusions and to encourage the wrong work habits. Realistic fact-finding and evaluation is a prerequisite for any learning. Social research should be one of the top priorities for the practical job of improving intergroup relations (p. 35).

Good managers want to learn; they want to promote the rational development of their activities. Far from diverting attention from project development, evaluation is a tool of project and organizational development (French & Bell, 1978). To use this tool, the proj-ect must be willing to grow and develop, and to devote resources to the learning enterprise.

Problems in implementing an evaluation often surface when a project has difficulty filling its institution or meeting service delivery quotas. Any kind of control over assignment--admissions, selection, differential treatment--is difficult when an organization has trouble at the front door. If its services are not in demand, or if the demand is for services of a different kind than the project aims to provide, problems are created for evaluation and for project implementation.

For example, teachers may refuse to make referrals to a treatment unit, or persons deemed in need of the services may not avail themselves of it. Then a problem exists with or without the evaluation: too few

eligible candidates for the project. This problem may affect only the evaluation (which is rarely the case); pools of eligibles will not be of sufficient size to make control over assignment possible. In this situation, the small number of candidates could indicate a lack of commitment to self-study and evaluation. On the other hand, the problem may affect other aspects of implementation. The unavailability of a pool of eligibles may result in the provision of services to an inappropriate group, or the project may be unable to attract clients truly in need of services. In this case, a more serious project management problem exists.

Ironically, the failure to take steps to evaluate rigorously may lead to the perpetuation of the problem because learning does not take place, interventions may not be modified to become more appropriate or attractive, and the organization has only soft evidence or vague appeals to use in its effort to extend services to persons deemed in need of them. At the very least, the unavailability of sufficient pools of eligibles in a pilot study or demonstration project is a major defect in a project of this type, because it limits what can be learned about the effects of the project.

Evaluation Is the Problem

"Elements of the evaluation procedures thwart the project because the organizational hierarchy, the project's staff, or other elements in the project's environment resist it; this threatens the project's stability."

As Day (1981) has pointed out, a project has a higher likelihood of being institutionalized and of creating change in the system if certain conditions are present. Among these are the following: (a) Key decisionmakers have a reputation for innovation and experimentation. (b) The system is monitoring the project and receiving information about it. (c) The project sees itself as a demonstration or pilot project.

A static organization, one that does not wish to create change or one that clings to the status quo, is naturally wary of rocking the boat. Beer (1980) notes two major sources of change in an organizational system: crisis and information. Sooner or later, an organization that avoids the impetus to change provided by information may be expected to experience crisis.

Organizational development specialists generally believe that the persons affected by a project or by research on it should be involved in the development of research questions and designs (see e.g., Frohman, Sashkin, & Kavanaugh, 1976).

This involvement, which we aim to foster in the School Action Effectiveness Study, is important not only in overcoming resistance to evaluation, but also in fostering the subsequent utilization of information generated by the evaluation. A useful strategy for a project director experiencing staff resistance would be to replicate parts of the Program Development Evaluation process with members of the project's staff or the organization's hierarchy. This approach may encourage the members' support for activities they would come to see as relevant and important.

Evaluation Restricts Freedom

"My project is trying to get started. We have enough problems without trying to adhere to onerous structures created by an evaluation."

When a project is going through the first months of starting its operation, activities aimed at discovering the effectiveness of its interventions are usually inappropriate. This is especially true when procedures have yet to be developed, staff members are still unclear about project goals or perhaps have not yet been hired at all, space has not been rented, and no services are being rendered. At this point, evaluation activity is appropriately aimed at clarification of project goals, rationale, objectives, plans, history, implementation, and setting.

Even at this early stage, however, it is essential to begin planning for summative kinds of evaluation activities. This is especially true with time-limited demonstration projects. If early steps to develop a framework for evaluation are not taken, the project may never be evaluated in its lifetime. The trick is to balance activities so that they are appropriate for the developmental stage of each project at any point in time.

The possibility exists that some projects will remain in what is essentially a start-up stage for several years. In such a case, evaluation might appropriately be limited to a process evaluation for the entire lifetime of the project, focusing on such issues as the most effective way of recruiting and training staff, or the best means of developing referral pools and initial intervention methods. Inferences about project effectiveness are only appropriate after some plausible interventions have been fully implemented.

The Taxpayer Over Our Shoulders

In this Alternative Education Program, the emphasis on evaluation has a basis that goes beyond its utility as a sound management practice. The OJJDP has awarded grants to the alternative education action projects involved in SAES totalling $10,944,442. Congress and the Office of Management and

Budget willing, some of these grants may be supplemented in future years--even more public funds will be spent. In addition, OJJDP is spending over $300,000 of taxpayers' money for technical assistance and for evaluation each year. This expenditure of public funds is justified in large measure on the basis of the knowledge that will be gained about alternative education and delinquency prevention. There is no justification for expending these funds without conducting the most thorough and rigorous evaluation possible. As the background paper for the OJJDP Alternative Education Program announcement (1980) put it, "Without standardized measure, rigorous evaluation designs, and adequate follow-up time frames, we will continue to be unable to assess the effectiveness of alternative education for delinquency prevention. Policy and funding decisions will continue to be made without such knowledge" (p. 43). Therefore, the Program Announcement required all applicants to "provide assurances in their application(s) agreeing to cooperate with the national evaluators in terms of . . . the overall evaluation component" (p. 9).

Implications

Scientific, practical, moral, and programmatic considerations converge in demanding the most rigorous possible evaluation of the activities undertaken as part of the Alternative Education Program. If the projects and their umbrella organizations view their activities as demonstration efforts in an area where knowledge of what works is desperately needed, they must consider evaluation an integral and helpful aspect of project development.

To accomplish a rigorous evaluation, collaboration between project implementers and evaluators is essential. A structure designed to facilitate this collaborative process, Program Development Evaluation, is described in a subsequent chapter. The evaluation process has required, and will continue to require, effort and resources on the part of each action project. This is not surprising; good evaluations are costly, time consuming, and demanding.

References

American Psychological Association Task Force on Continuing Evaluation in National Health Insurance. Continuing evaluation and accountability controls for a national health insurance program. American Psychologist, 1978, 33, 305-313.

Bachman, J. G., Green, S., & Wirtanen, I. D. Dropping out--Problem or symptom? Ann Arbor: Institute for Social Research, 1971.

Beer, M. Organization change and development: A systems view. Santa Monica: Goodyear, 1980.

Boruch, R. F. On common contentions about randomized field experiments. In R. F. Boruch & H. W. Riecken (eds.), Experimental testing of public policy. Boulder, Colorado: Westview, 1975. (Reprinted in G. V. Glass (ed.), Evaluation studies review annual (Vol. 1). Beverly Hills: Sage, 1976.)

Cook, T. D., & Campbell, D. T. Quasi-experimentation: Design & analysis issues for field settings. Chicago: Rand McNally, 1979.

Day, N. Indicators of commitment to institutionalization or system change, and indicators of potential for public schools to institutionalize project or adopt key elements. Remarks at the Alternative Education Program Technical Assistance and Evaluation Workshop, Baltimore, Md., and Washington, D.C., August 1981.

Eisenberg, L. The social imperatives of medical research. Science, 1977, 198, 1105-1110.

Elliott, D. S. Delinquency, school attendance and dropout. Social Problems, 1966, 13, 307-314.

Feinberg, S. E. Randomization and social affairs: The 1970 draft lottery. Science, 1971, 171, 255-261.

French, W. L., & Bell, C. H., Jr. Organization development (Second Edition). Englewood Cliffs: Prentice-Hall, 1978.

Frohman, M. A., Sashkin, M., & Kavanaugh, M. J. Action research as applied to organization development. Organization and Administrative Sciences, 1976, 7, 129-261.

Glaser, D. Strategic criminal justice planning (DHEW No. ADM 75-195). Washington, D. C.: U.S. Government Printing Office, 1975.

Gottfredson, D. M. Research--Who needs it? Crime and Delinquency, 1971, 8, 11-22.

Gottfredson, G. D. Evaluating vocational interventions. Journal of Vocational Behavior, 1978, 13, 252-254.

Gottfredson, G. D. Schooling and delinquency. In S. E. Martin, L. B. Sechrest, & R. Redner (Eds.), New directions in the rehabilitation of criminal offenders. Washington, D. C.: National Academy Press, 1981.

Hindelang, M., Hirschi, T., & Weis, J. G. Measuring delinquency. Beverly Hills, Calif.: Sage, 1981.

Lewin, K. Action research and minority problems. Journal of Social Issues, 1946, 2, 34-46.

Office of Juvenile Justice and Delinquency Prevention. Program announcement: Prevention of delinquency through alternative education. Washington, D.C.: Author, 1980.

Park, J. W. L. A tale of two tigers. Correctional Review, 1965 (July-August), 3-7.

Tharp, R. G., & Gallimore, R. The ecology of program research and development: A model of evaluation succession. Unpublished manuscript. (Available from Roland Tharp, Department of Psychology, University of Hawaii, Honolulu 96822, or from Ronald Gallimore, Department of Psychiatry and Biobehavioral Sciences, UCLA, Westwood, California 90024. (A secondary account of the evaluation succession approach is available in I. McNett, KEEP early education project . . . from research to evaluation to success, APA Monitor, December, 1981, pp. 8-9, 33.)

Footnotes

1. Throughout this chapter "treatment" is used as a shorthand description of any intervention intended to bring about an effect in a person, group, or school. It neither specifies the nature of the outcome intended, nor implies any particular modality of intervention.

2. Each of the other possibilities mentioned has been explored at one time or another with at least one of the alternative education projects.

3. Randomization may not be the evaluator's method of choice if a project is still in the stages of developing its intervention in potent form, if it is floundering in developing any intervention at all, or if the obstacles to randomization are such that the intervention itself is subverted by experimentation. Tharp and Gallimore (n.d.) describe stages in project development where methods other than true experimentation may be most productive. But randomization is generally the method of choice when a project can implement some plausible intervention with fidelity, when it can continue to function if randomization is present, and when doubt about the intervention's effects exists.

4. This story came to our attention when retold by Gottfredson (1971).

Program Development Evaluation

The School Action Effectiveness Study (SAES) is faced with a tough, but not unusual challenge. Evaluations--not only delinquency prevention evaluations--are commonly marked by weak interventions, or interventions of unknown strength and integrity; a lack of theory; a rapidly changing project environment and changes in project goals, objectives, or methods; little commitment of project implementers to evaluation, and little understanding on the part of evaluators of the problems of implementation; fears about the ways evaluation results may be used; a lack of sound measurement of the outcomes of interest; weak or nonexistent evaluation designs; ambiguities about goals, objectives, problems, and needs; and inadequate resources (time and money) to cope effectively with all these problems.

The history of previous delinquency prevention evaluations, described by Ogawa (this volume), is a history of evaluators and project implementers grappling with these problems. The success of SAES will depend upon the extent to which it develops methods that avoid these difficulties. The development of such methods has an importance that goes beyond the Alternative Education Program. Evaluators and program developers everywhere need practical and sound methods for improving project implementation and fostering more useful and rigorous evaluations.

To meet its challenge, SAES must implement an evaluation structure to meet the following demands:

1. Increase the likelihood that plausible and potent interventions will be implemented.

2. Make possible the assessment of the strength and fidelity with which interventions are implemented.

3. Provide for the interpretation of experience in theoretical terms.

4. Document project plans and their implementation as they evolve, recognizing that in actuality plans neither stay put nor are necessarily followed.

5. Conduct the most rigorous evaluation possible in terms of the strength and relevance of the design and the measurement of key outcome variables.

I am grateful for the advice of Deborah Daniels, Denise C. Gottfredson, and Jane St. John on a draft of this chapter, and for the discussions with J. Douglas Grant and Carol Yamasaki in the early stages of the creation of the Program Development Evaluation model.

6. Do all this with limited resources in a short period of time.

As is typical of many evaluations, the evaluator is an "outsider." That is, each action project competed independently for funding on the basis of the strength of its proposal. Evaluatability was not a major consideration in the funding decisions. Action projects are not under the control of the evaluator, nor did the evaluator have any hand in the selection of action projects. Under these circumstances, action project personnel may perceive evaluation as something imposed upon them by an alien and perhaps unfriendly agent. Avoiding this perception is important, because we expect that projects will increase in effectiveness over time in proportion to their use of evidence provided by the evaluation.

Instrumental in meeting these six demands, therefore, is a further demand that SAES gain the cooperation of action project implementers. This is essential to (a) increase the extent to which the evaluation is directed to the aims of each action project rather than focusing solely on the goals of the Federal sponsor, (b) increase the extent to which evaluation methods and results are used by the action project in its development, (c) ensure that action projects devote an appropriate level of resources and attention to evaluation, (d) increase the rigor and relevance of the evaluation, (e) assist in clarifying goals, objectives, and plans to focus the evaluation and to assist in project development, and (f) capture the rationale or theoretical perspective of each project, directing attention to the implications of these perspectives for the development of delinquency prevention efforts.

Implicit in this list of demands is the need for an effective evaluation to go beyond the two approaches common in many evaluations. Some evaluators approach their task in a wooden way, imposing a common set of measurement and design requirements that may be insensitive to the aims or circumstances of the action projects, and that are likely to be passively or even actively resisted by project implementers. This approach seldom fosters project development. The evaluation may end up hopelessly corrupted or may assess a set of interventions that were never implemented as intended.

A second common approach is to conduct a flabby evaluation. Unfortunately, a frequent response of evaluators to the six tough challenges is to abandon rigor. Evaluations are often limited to attempts to obtain flow data; efforts to implement procedures leading to confident conclusions about effectiveness and to assist

in project development are abandoned. Weak evaluation designs often result in an inappropriate level of optimism, with little evidence of effectiveness after lots of taxpayers' dollars have been spent. SAES aims to conduct the most rigorous possible evaluation while being flexible and useful to project implementers.

The diverse demonstration projects involved in the Alternative Education Program aim to alter organizational forms and educational experiences to prevent delinquency. The common goals and objectives of the program, specified by OJJDP (1980), form a core or common basis for the evaluation of all the projects in the national program. But projects are run by community-based organizations, school systems, and a university, and each has distinctive problems and goals. Overlaid, therefore, on this common framework is a diverse set of organizational environments, goals, objectives, and interventions specific to each project.

A Program Development Evaluation (PDE) model provides the structure for the evaluation of these projects. The structure is intended to anticipate and foster the development of these projects by involving project personnel in a cycle of evaluation activities. This structure is intended to (a) make rigorous evaluation possible, (b) make the evaluation relevant not only to national concerns but also to the concerns of project personnel and managers, (c) document project implementation, (d) facilitate project implementation, (e) tie the evaluation explicitly to delinquency prevention theory, and (f) integrate research with project operations so that projects develop by using the results of research in project planning. Related structures, differing somewhat in emphasis and detail, are provided by Empey (1980) and Tharp and Gallimore (n.d.). Those related structures are guided by some of the same concerns that led us to develop the PDE structure.

The Program Development Evaluation model stems from the action research model. This approach assumes that the prospects for promoting change are greatest when the program decisionmakers' stake in the research is made clear by their own participation in the research. Project decisionmakers and researchers collaborate through a continuing dialogue in which researchers provide feedback on the consequences of project action. Action research involves a cycle of hypothesis formulation and planning, action, evaluation and information feedback, and then renewed hypothesis formulation and planning. As the cycle is repeated, and information derived from project efforts and research is used in decision making, projects should become more effective--turning the process into an upward spiral of activity.

Projects usually change over time on the basis of the experience gained as they develop (Wilkins & Gottfredson, 1969). What Pearl (1962) has called "quality control" is needed to insure not just that a program is run according to the plan, but that a plan exists and is modified to coincide with the way a project, as it develops, is actually run. Many attempts to demonstrate the effectiveness of specifiable social programs have failed in part because plausible interventions were not implemented or their implementation was not documented (Quay, 1977; Sechrest, White, & Brown, 1979; Hall & Loucks, 1977), or the plan for the innovation was not clearly articulated at the outset (Sarason, 1971).

The PDE model is especially well-suited for facilitating and studying the development of a program by assisting in the planning process. It provides a mechanism by which an organization can make its plans explicit, and then engage in self-study as it goes about implementing them. It also helps the evaluator monitor and document project plans and their implementation as the project evolves. In short, our Program Development Evaluation is an attempt to integrate evaluation and organizational development. Its action research approach to knowledge generation and organizational growth is derived from a tradition of concern for practical theory, useful research, and organizational change and development.

Antecedents of PDE

Organization Development

One of the roots of Program Development Evaluation is the practice of organization development (OD). French and Bell (1978) characterize OD as a process involving action research that emphasizes normative change, is based in behavioral science, involves experience-based learning of intact work teams, and emphasizes goals and objectives. By characterizing OD as a process, French and Bell mean that OD is "not to be regarded as a one-shot solution to organizational problems, but more as a 'growing toward' greater effectiveness through a series of intervention activities over a period of time. . . . Changing the culture of . . . an entire organization is a long-term, involved process (p. 69)." In addition, they see OD as a process involving rational, empirical strategies, but one that is even more dependent on normative-reeducative strategies: "The client defines what changes and improvements he or she wants to make, rather than the change agent; the change agent attempts to intervene in a mutual, collaborative way with the client as they together define problems and seek solutions; anything hindering effective problem

solving is brought to light and publicly examined (pp. 75-76)." The emphasis on normative education is based on the assumption that behaviors are rooted in norms, values, or beliefs as well as in rationality and self-interest. OD is a data-based approach to planned change in which information is a spur to action. Unpleasant information is not to be avoided but rather treasured because it may lead to advancement, to clarification of problems. Typically, OD emphasizes concrete goal setting through the shared experience of a group in formulating plans. The on-the-job learning experience of an intact group is presumed to promote organizational and individual effectiveness.

The interactive, collaborative, participative approach, often used by behavioral scientists or OD specialists serving as consultants or facilitators of organizational planning and decision making, has much to offer in overcoming some of the difficulties an evaluation may expect to face. First, increasing an organization's effectiveness should increase the likelihood that it will succeed in implementing interventions with a possibility of being shown to be effective when subjected to serious summative evaluation. Second, in the OD process, the evaluator approaches an organization in a manner that may decrease the extent to which he or she is perceived as an alien invader. By helping an organization clarify its goals and objectives, by assisting in creating open communication about problems, and by fostering the expectation that projects will change and develop over time, the evaluator may come to be considered more as an insider, an entity to be trusted to convey useful news. And, the perspective that information, even uncomfortable information, is valuable in fostering growth and confronting important problems may decrease the organization's usual fear of evaluation. Finally, the links between OD and action research make the interjection of formal evaluation possible.

The Program Development Evaluation model is in part a descendent of an OD method previously used by the Social Action Research Center (Blanton & Alley, 1975) in a series of projects to manage and study social change. This predecessor, called the Program Development (PD) model, was developed through attempts to evaluate human service projects. In this model, feedback is a mechanism of project development that involves monitoring a project's environment, the implementation of strategies, and the achievement of goals. In practice, the Program Development specialist focuses on interaction with project implementers to assist in assessing needs, in articulating goals and more specific objectives, in analyzing a project's forcefield (environmental constraints and resources), and in developing strategies for change or implementation. Blanton and Alley (1975, Chap. 7) distinguish three kinds of evaluation possible using PD concepts: (a) evaluation of relevance, (b) procedural evaluation, and (c) outcome evaluation. Although they discuss potential structures for outcome evaluation, emphasis has been placed primarily on other aspects of the organization development process. The PD model is illustrated in Figure 1. In application, great emphasis has been put on the participatory nature of this process and on avoiding intrusive monitoring procedures to enhance the credibility of evaluation designs. Participation and unobtrusiveness facilitate the implementation of the planning portions of the PD process, and reduce the possibility that PD will become an unwelcome or burdensome appendage. Like other forms of OD, PD emphasizes participatory planning in part to foster normative-reeducation and in part to increase organizational and individual competencies in decision making and planning.

Action Research

Both OD and the present evaluation have roots in action research. According to French and Bell (1978), the origins of action research lie in the work of Dewey (1933), Collier (1945), and Lewin (1946). The roots of action research, however, are deeper than this. They can be traced back to the Baconian formulation of the scientific method, which specified three steps: (a) the formation of hypotheses, (b) the empirical testing of the hypotheses, and (c) the acceptance or rejection of the hypotheses (Deese, 1972). Action is taken to "twist the lion's tail" to learn about nature. Since Bacon, science has been active rather than speculative, historical, or reflective. Dewey translated the scientific method of problem solving for laypersons, and Collier and Lewin both applied the scientific method to solving practical social problems.

Collier, a commissioner of Indian Affairs concerned with improving race relations, wrote of action research, claiming that:

> Research and then more research is essential to the program, that in the ethnic field research can be made a tool of action essential to all other tools, indeed that it ought to be the master tool. But we had in mind a particular kind of research, or, if you will, particular conditions. We had in mind research impelled from central areas of needed action. And since action is by nature not only specialized but also integrative. . ., our needed research must be of the integrative sort. Again, since the findings of the research

> must be criticized by them through their experience, the administrator and the layman must themselves participate creatively in the research, impelled as it is from their own area of need.
> (Collier, 1945, cited by French & Bell, 1978, p. 94).

Broader attention was called to action research by Lewin, an eminent and influential psychological theorist with a keen interest in the applications of psychology. He saw that cooperation between the change agent (or field worker) and the researcher is important for both planning and management:

> Planning starts usually with something like a general idea. For one reason or another it seems desirable to reach a certain objective. . . . The first step, then, is to examine the idea carefully in the light of the means available. Frequently more fact-finding about the situation is required. If the first period of planning is successful, two items emerge: an 'overall plan' of how to reach the objective and a decision in regard to the first step of the action. Usually this planning has also somewhat modified the original idea. The next period is devoted to executing the first step of the overall plan . . . [and] by certain fact-findings. . . . This . . . fact-finding has four functions. It should evaluate the action by showing whether what has been achieved is above or below expectation. It should serve as a basis for correctly planning the next step, [for] modifying the 'overall plan.' Finally, it gives the planners a chance to learn, that is, to gather new general insight . . . regarding the strength or weakness of certain . . . techniques of action. . . .
>
> Rational social management, therefore, proceeds in a spiral of steps each of which is composed of a circle of planning, action, and fact-finding about the result of the action.
> (Lewin, 1947, pp 333-334).

This sequential and spiraling model of problem solving is now widely used in organizational development efforts, and has been applied in a variety of industrial, human service, and educational action research projects; and it appears to be at the heart of Tharp and Gallimore's (n.d.) Evaluation Succession model.

Several varieties of activity are often called action research (Chein, Cook, & Harding, 1948). Sometimes the effort is limited to diagnosis and recommendations; sometimes

organizations or project implementers carry out the entire process; sometimes records or diaries of actions taken and their perceived effects are maintained. As Tharp and Gallimore (n.d.) note, there are several ways of "knowing," each appropriate to different stages in the development of a program. What they call "experimentation," "qualitative/personal knowing," "data guidance," and "program evaluation" are all useful in program development and evaluation. But the variety of action research most productive of trustworthy knowledge is experimental action research. Unfortunately, experimental action research is also the most difficult to perform, because it requires the conditions necessary for confident inference (see the preceding chapter), and a stable set of interventions that the organization knows how to and can implement in testable form. Seldom do OD efforts aim to implement experimental action research, largely because it is so difficult. Implementing experimental action research is, however, a chief aim of the School Action Effectiveness Study. Evaluation must be coordinate rather than subordinate to problem solving; solving problems without learning how or why they were solved will not accomplish the aims of the Alternative Education Program or of SAES.

The PDE Structure

To conduct the School Action Effectiveness Study, we have built on the Program Development model, but have altered it in major ways to make this OD tool more appropriate for an evaluation. In particular, the PDE model emphasizes to a far greater extent (a) theory, (b) measurement, and (c) experimental or quasi-experimental design. In addition, some terms (most notably "objectives") have been redefined, and a structure for documenting project implementation has been added. At the same time, the new structure retains the action research emphasis on a cycle of development activity that was central to PD.

The resulting Program Development Evaluation model, illustrated in Figure 2, thus incorporates theory as an explicit component, gives measureable goals and objectives a more hard-nosed meaning, incorporates planning for evaluation implementation in the same way that planning for any other aspect of a project is incorporated, and allows project implementers and evaluators to monitor critical benchmarks in the implementation of any strategy to create change. The principal concepts involved in the PDE structure are listed in Table 1, and each is elaborated below.

PDE Concepts

The Alternative Education Program involves a common set of goals; directed primarily at the problems of delinquency, dropout, and nonattendance in school. But a fundamental tenet of the action research paradigm is that the implementers of an individual project should be actively involved in creating the research project and setting its goals. Furthermore, seldom do the aims of any particular action project overlap fully with the aims of the sponsor of this program. Therefore, interaction with each project must begin with an exploration of its intent.

Problems and Goals. Most organizations, and most projects, have multiple aims. Within the PDE framework, a general or overarching aim is called a goal. A goal is the obverse of a problem; it specifies how the level of the problem may be measured and therefore how one may know if progress is being made. Several secondary questions are important when discussing goals. The first question--how each goal may be measured--serves to reduce ambiguity and enable evaluation. The second question serves to promote realistic planning; it asks when a project can realistically expect to make a substantial difference. And the third question, essential in experimental or quasi-experimental action research, asks how one may know that the project itself was responsible for progress towards the goal. These questions are, of course, steps toward involving project implementers in the design of the evaluation.

Theory. Actions are taken for reasons that are either articulated or unarticulated. The PDE structure is a vehicle for making theory explicit. This is useful because, as the Panel on Research on Rehabilitative Techniques (Martin, Sechrest, & Redner, 1981) notes:

> In attempting to solve any problem, a clear idea of the nature of the problem, its causes, and developmental processes is vital. In the absence of an adequate conceptual framework . . . the rush of enthusiasm for an interesting intervention is likely to short-circuit consideration of these factors. The result is . . . efforts that may be unrelated to the causes of crime, ignore the most suitable target populations, and fail to consider questions of optimal timing and strength of the intervention. The adoption of a theoretical framework necessarily prompts consideration of the above factors and, one hopes, thoughtful development and implementation of . . . interventions, thereby increasing the chances for effectiveness (p. 29).

Theory helps to organize knowledge and to communicate, it provides a guide for action, and it assists in developing and assessing interventions. "Once a basic problem is stated in theoretical terms, planners have an explicit foundation on which to build an intervention strategy and from which to derive a research strategy in conjunction with the intervention" (Martin et al., 1981, p. 34; cf. Glaser, 1980). In short, an explicit theory provides a template for project implementers' use in building their interventions, as well as a template by which both implementers and researchers can assess those interventions. Therefore, the PDE process calls for deliberate and careful consideration of the question, "Why do these problems exist?"

Objectives. In the language of PDE, an objective is an intermediary outcome that a project's theory of action implies is important.<1> Like goals, objectives must be stated in measurable terms.

Some examples may help make the distinction between goals and objectives clear. Suppose that a change agent wishes to decrease the death rate due to gastroenteritis in a rural society. The change agent theorizes that the suffering and death are due to the contamination of village water supplies with the cholera micro-organism. This theory might suggest a campaign to chlorinate wells, with the objective of decreasing this contamination. The objective would be measured by laboratory analyses of well-water samples to determine the levels of microbial contamination, and the goal might be measured by counts of deaths per 100,000 population due to gastroenteritis. Another change agent might see the problem somewhat differently. This second change agent may theorize that the suffering and death are due to poor environmental sanitation: Because few villagers use sanitary latrines, well water is easily contaminated and the cholera micro-organism spreads from infected to uninfected persons. This theory might suggest an environmental sanitation campaign directed at persuading villagers to construct sanitary latrines and sanitary wells. The objective now involves villager behavior, and might be measured by the proportion of households using sanitary latrines and water from protected wells. A theory can, of course, suggest multiple interventions and multiple objectives. The second change agent's theory would also reasonably imply chlorination of wells and assessments of well water. The more comprehensive a theory, the more complex the array of interventions and objectives it is likely to suggest.

Change agents could develop theories at many levels to explain the problem of cholera deaths, and each level would suggest

somewhat different interventions. To continue the examples, change agents might attribute the problem to (a) normative beliefs in village societies that current standards of environmental sanitation are adequate, (b) the poverty and segregation of the rural people, which deprive them of the resources to build sanitary devices and concentrates them so that they are at high risk, (c) social stratification that allows only an elite merchant class access to sufficient resources to enjoy a sanitary environment, (d) stratification in the world system that enables capitalist countries to keep countries with rural rubber-tapping populations impoverished and the cost of raw materials low. Each of these theories may have considerable validity. Yet each would imply different interventions to solve the problem, ranging from dumping chlorine in wells to overthrowing the capitalist world system. No single cholera prevention project is likely to attempt interventions at all of these levels, and so will not have objectives at each level. A project's theory of action--the theory that drives its interventions--is the theory that is relevant in specifying objectives.

Again, answers to questions--how objectives may be measured, when effects are to be expected, and how one may know that the intervention caused the effects--serve to create the evaluation design.

Intervention. An intervention is an action taken to achieve an objective or set of objectives. Ordinarily, it is a major component of a project. The term is often synonymous with "change," "treatment," or "component." Some interventions are aimed at changing the behavior, attitudes, or status of individual people; others are aimed at changing the behavior of an organization or collectivity.<2> An intervention is a process, action, structure, rule, or substance that a project applies or puts in place to achieve an objective or set of objectives, and therefore to move closer to achieving its goal(s). An intervention may be chemical, physical, biological, behavioral, social, political, or structural.

Forcefield. A forcefield is the social-psychological field that immediately surrounds a decision or action. It includes the forces that compel or restrain against alternative actions, as they are perceived by an individual or corporate actor. The notion of a forcefield comes from Lewin's (1951) ideas about the field of forces influencing action. An examination or analysis of an organization's forcefield, especially one that focuses on the field in terms of the resources available and the obstacles to action, is frequently useful for four reasons: (a) By focusing on the organization's perceptions of environmental influences, the nature of these perceptions becomes explicit and open to scrutiny, revision, ammendation, supplementation, and test. (b) A complete account of obstacles and resources decreases the likelihood that either pitfalls or potentials will be overlooked in the development of a project. (c) Using knowledge of the influences in the project's environment helps to capitalize on opportunities or arrangements that go beyond the resources under a project's direct control. (d) Alternative strategies or plans to implement any intervention can be created and assessed in the context of the forcefield.<3>

Practical guidance on working with an organization to analyze its forcefield is provided by Blanton & Alley (1975, 103-113).

Because initial analyses of a forcefield may be objectively incorrect, because perceptions change over time, and because the action of a project may alter its forcefield, the dynamic nature of the field is to be expected. A sensible practice, therfore, is to renew forcefield analysis periodically, especially when any strategy being executed on the basis of an initial forcefield analysis is not working well.

Strategies. Strategies are plans.<4> According to the PDE model, strategies are developed from a forcefield analysis, just as objectives and interventions derive from a theory of action about a problem.

Several possible strategies for implementing a project or one of its component interventions are likely to exist. The task for project implementers and those who are attempting to facilitate strategy development is to create a plan that is perceived as feasible and attractive. If a critical path in some plan is blocked and no way around the obstacle is perceived, the plan is not a good one. Alternative paths that objectively exist but have not been perceived will not be followed. (This point illustrates why thorough and creative forcefield analysis is helpful.) A strategy that appears workable will make use of an organization's resources to overcome the obstacles to implementation. Such a strategy may involve (a) moving around an obstacle, (b) decreasing the strength of the forces working against implementation, (c) turning an obstacle into a resource, or (d) involve a strategy in which the obstacle is irrelevant and need not be overcome.

A fully articulated strategy is composed of two kinds of elements: critical benchmarks and tasks.

Critical benchmarks. A critical benchmark is a key decision, agreement, action, or arrangement necessary to move forward with a plan. A benchmark is much like a gate that must be opened to move along a path.<5> If the gate does not open, progress in executing the strategy is blocked. The locations of

these benchmarks (or the nature of them) are made clear in the process of analyzing the forcefield around an intervention. For example, the forcefield analysis about a project's efforts to provide in-service training for teachers might imply that an obstacle lies in teacher unwillingness to participate in training outside of normal working hours, and that a resource is the authority of the deputy superintendent of schools to grant release time and to allocate the funds for substitute instructors. The deputy superintendent's agreement to grant release time and to authorize the expenditure for substitute teachers would then become a critical benchmark. The deputy superintendent is a gatekeeper (Lewin, 1947, p. 333) whose psychology must be examined to learn how to get the gate opened.

Specifying when a critical benchmark is to be accomplished provides a management tool. Any strategy will require a temporal or logical sequence of milestones that must be met. In this example, a failure to accomplish this critical benchmark would signal the need to devise a new strategy for getting the training done, or the need to seek an alternative to training.

Tasks. The second part of a strategy is the collection of tasks required to execute it. A task statement specificies who will do what by when.<6> Specifying a person to be responsible for executing a particular task, even when a group will be involved, promotes clarity. And specifying when a task is expected to be completed is an additional management tool.

Critical benchmarks and tasks both serve important functions in project management and worker reinforcement: They serve to guide an organization's efforts. They provide one kind of objective standard of achievement. A lack of such objective standards "deprives the workers . . . of their legitimate desire for satisfaction on a realistic basis. Under these circumstances, satisfaction or dissatisfaction with (one's) own achievement becomes mainly a question of temperament (Lewin, 1946, p. 35)."

Development. At the very heart of the PDE model is the expectation that project development will be an ongoing process, and that the project's environment is dynamic. Only an effete organization is immobile, at equilibrium. Tension, reassessment, review, replanning, and changes in actions taken are the hallmarks of vigorous projects. Consequently, PDE is a cyclical process of action research as progress is made towards achieving goals and objectives (or as goals and objectives are redefined), as new information becomes available, and as the environment changes.

Development occurs largely through the use of information. Information about the achievement or nonachievement of critical benchmarks signals that the forcefield has been usefully understood, or that developmental effort is required to reassess the organization's forcefield. Information that an objective is being achieved signals that an intervention is effective, and information that an objective is not being achieved signals a reconsideration of the appropriateness, strength, or fidelity of the intervention, and prompts new planning. Information that there is progress towards a goal signals that the organization is on the right track. Information that there is no progress towards the goal may signal several things, depending on the pattern of other feedback. If interventions are being implemented as intended and they are achieving their objectives, the theory is called into question. If objectives are not being met, either the theory or integrity of the intervention, or both, should be scrutinized. Success in bringing about elusive objectives and solving serious problems is not to be expected at once. But the PDE structure is intended to provide interim feedback on progress to enable a strengthening of the project.

Evaluation. The PDE structure is intended to facilitate several kinds of evaluation. The explication of a theory of action allows an assessment of its plausibility, and an assessment of the plausibility or strength of the project's planned interventions in light of the theory.

By tracking the achievement of critical benchmarks, the structure allows assessment of the integrity with which an intervention is executed--it provides evaluators and project managers with feedback on progress towards executing strategies. These are key elements of formative evaluatin.

The PDE structure is also intended to facilitate rigorous summative evaluation--it is experimental action research (Chein, Cook, & Harding, 1948) or at least quasi-experimental action research. It repeatedly asks the question, "How do you know your intervention (project) made the difference?" The implementation of an evaluation design is treated in the same way as the implementation of any other intervention. Essentially, the PDE model assumes that evaluation is an essential component of effective project development and should receive coordinate effort with other aspects of project implementation. Therefore, forcefield analysis is performed for design and data collection issues just as it is for any other project component. Because project implementers are involved in the research design and in the specification of the

research questions, their commitment to strong evaluation is expected to increase. And, because the forcefield analysis focuses on the project implementers' own forcefield--their perceptions of the possible--the immediate environment of the evaluation is taken into account when the evaluation is designed, perhaps mitigating some of the resistance to evaluation activities commonly encountered among implementers.

Limitations and Potential Criticisms

The PDE structure in its current state of development has some limitations, creates some tensions, and is open to criticism. The most important appear to be that it is complex, it is time and expertise intensive, it does not yet directly assess the strength and fidelity of interventions, it fails to completely resolve the tensions summative evaluation causes for project implementers, it is an imperfect mechanism for coping with the separate goals of project sponsors and implementing organizations when these are not completely in accord, and it confronts researchers and implementers with tough decisions involving the sacrifice of rigorous research designs in order to achieve some aspect of project implementation.

Time and Talent

The human behavior required to successfully implement the PDE model is complex, and the model's implementation calls for a large investment of human resources. Use of the PDE structure calls for high levels of interpersonal competency, tact, patience, communication skill, and understanding of group relations in organizations. In addition, it calls for a thorough understanding of evaluation methods--measurement, social science theory, experimental and quasi-experimental methods, statistics, and rigid adherence to schedules. Ironically, this combination of competencies are rarely found in one and the same person, suggesting that a team of workers may be required to conduct action evaluation using PDE. Furthermore, the cyclical or developmental natures of PDE requires constant (or at least frequent periodic) attention, monitoring, updating, and information communication. This intensive and taxing process goes beyond the effort typically expended in an evaluation.

Some trade-offs are likely to be involved in staffing an evaluation using the PDE model. Finding staff members with the requisite skills can be difficult, implying that training will be required. Our experience implies that para-professionals can function effectively using the model, but that they will require assistance with the

more technical evaluation and statistical issues. It also implies that social scientists trained primarily in research methodology, statistics, and theory can successfully implement the model, but that they require a different kind of additional training to do so. By using training to develop competencies, and a staff composed of persons with diverse skills, personnel costs may be kept relatively low despite the scarcity of persons with all the competencies required.

Strength and Fidelity

The PDE model makes possible the assessment of strength and fidelity through judgments about theoretical plausibility and benchmark monitoring. This assessment can occur in two ways. First, project implementers can assess the consistency of their interventions and objectives with the theory of action underlying their project. That is, a project implementer can determine whether the objectives sought accord with the theory, and whether the interventions planned will plausibly achieve the project's objectives. In short, theory is a template for making judgments about the appropriateness of interventions and objectives that project implementers can use to quality control their own projects. Second, observers of a project can assess its <u>a priori</u> strength by determining whether the planned interventions will plausibly lead to the objectives or goals of the project by assessing them in comparison to state-of-the art theories in the field and the history of similar projects that have been conducted in the past.

In implementing PDE in the past, we have not, however, typically attempted to observe directly the administration of treatments or the conduct of interventions. As the utility of "manipulation checks" in experimental social psychology implies, the direct assessment of treatments would be desirable. The Tharp and Gallimore (n.d.) account of evaluation succession implies that monitoring the integrity of interventions is more important in some stages of the development of a project than in others. At some points in a project's development, evaluation issues will have to do with ideas for interventions or with strategies for getting an innovation adopted. At other points, evaluation issues will have to do with the integrity of the intervention's implementation and with the assessment of effectiveness.

Accordingly, we plan to pay more attention to the documentation of interventions as implemented where appropriate in the future. This documentation may take the form of detailed manuals for the administration of treatments or programs; descriptions of the characteristics of staff and target groups; and accounts of

the duration and scheduling of treatments or events, treatment protocols, or proportion of the population served. In many cases direct observation or mini-ethnographies may be helpful.

Some shift in emphasis in the application of the PDE model is desirable, because in the first year of this evaluation we have focused primarily on the steps leading to the adoption of a practice or innovation rather on the steps leading to the integrity of the innovation once "adopted." One minor shift in emphasis to increase the ability of the PDE model to assess integrity is to give more emphasis to those aspects of project planning that are aimed at achieving integrity of an intervention. For example, PDE plans could easily include strategies to develop manuals to guide service delivery, the making of diagnoses, and the training of staff. Similarly, PDE plans could include strategies to monitor staff performance, provide incentive structures to keep performance according to specifications, and the like. The implementation of those strategies would likely have two consequences: (a) the plans and their execution would <u>increase</u> the integrity of the intervention, and (b) the information generated by the implementation of these plans would <u>describe</u> the integrity of interventions.

<u>Tensions</u>

Tension appears endemic in summative evaluation efforts. Too often in the past, evaluation has been used as a tool for cancelling a project--even when positive evaluations could not reasonably be expected at an early stage of project development. Tension is also created by the inherently political environment of action projects, and by environments where the successful project does not rock the boat. Rigorous evaluation requires the expenditure of time and money, and often implies the necessity of arrangements that are disruptive.

<u>Divergent Goals</u>

Although the PDE structure is explicitly designed to focus on the goals and objectives of an action project, at present the goals and objectives of the sponsoring organization must be overlaid on these project-specific aims--and the overlap is sometimes imperfect. This is not a problem for an organization conducting its own development effort, but it creates resentment or resistance in some cases when a sponsor needs evaluation information that goes beyond what the implementing organization sees as relevant.

<u>Tradeoffs and Research Rigor</u>

Program development evaluation is value laden. Participation of project implementers is a fundamental

principle in the PDE process; pursuit of the goals and objectives of the implementing organization are generally assumed to be desirable (although open to question). Furthermore, an aim of PDE is to develop the implementing organization's capacity to accomplish its aims. Therefore, evaluators and implementers collaborate in evaluation design, question formulation, and planning. As a result, evaluators extensively intervene in project development--indeed they become a part of the project.

Some evaluators (Perloff, 1979) see this as undesirable in a summative evaluation because it raises questions about the generalizability of the results to situations where evaluators are absent. In addition, just as evaluation needs sometimes intrude in project operations by creating new tasks or structural arrangements, the pursuit of a project's programmatic activities very often results in compromises in research design. As Deutsch (1968, p. 466) says, "The danger that confronts the research worker in such situations is the possibility that his research design or methodology will be sacrificed to the achievement of the social-action objective."

This "danger" may account in part for the reluctance academic social scientists have shown to participate in action research. This danger seems a small price to pay in exchange for the opportunity action research creates to contribute to the solution of social problems, although the tradeoffs involving evaluation rigor are painful. In short, the PDE model is no panacea for this tough problem.

<u>Complexity</u>

The PDE model is complex. A comparison of this model with the PD model (Blanton & Alley, 1975) discussed earlier is sufficient to show that the increased emphasis on measurement, evaluation design, and theory, and the introduction of critical benchmarks, have resulted in a more cumbersome tool. Unfortunately, each component of the model seems at present to be useful and desirable in an effort such as the School Action Effectiveness Study. Nevertheless, this increased complexity suggests that a more streamlined model is appropriate when doing short-term organizational development interventions. Just learning the meaning of all the terms involved in the PDE structure is a large task. Consequently, for many brief organization development interventions, the selective use of those portions of the PDE structure that seem to be the most relevant for the problem at hand is more appropriate than attempting to use all parts of the model.

Practical Application

Summarizing experience with the use of the PDE model is difficult. Because the model was created for this evaluation, experience in its use is limited. Testimonial evidence suggests that one or another part of this process is useful to project implementers in defining their own jobs, in formulating plans, and in clarifying their intentions. Testimonial evidence also suggests that the entire process is sometimes viewed as burdensome. On the whole, this structure seems a clear improvement over some more traditional evaluations because it involves implementers in evaluation planning, because it explicitly attempts to build summative evaluation structures based on an organization's forcefield, and because it focuses on goals and objectives of concern to implementers.

The greatest virtues of the PDE model appear to be (a) its ability to elicit clear statements of the theory of action underlying a project, (b) its ability to elaborate clear measurable intervening outcomes, or objectives, useful in assessing the effectiveness of interventions, (c) its ability to provide project implementers with the tools to assess their own efforts by measuring interventions against theory and objectives, (d) its ability to generate strategies perceived as feasible to implementers based on the divergent thinking that takes place in forcefield analysis, (e) its ability to involve project implementers in the evaluation enterprise by engaging them in the specification of measurable goals and objectives, and in the creation of evaluation designs, (f) its ability to provide short-term assessments of progress through the monitoring of critical benchmarks and tasks, (g) its ability to enable evaluators to understand the nature of a project by translating implementer's ideas into a structured language of action research, and (h) its ability to serve as a structure for communication between researchers and practitioners that to some extent enables practitioners to become researchers by engaging in the study of their own actions.

To some degree PDE makes rigorous evaluation more attractive to implementing organizations despite its inability to make it truly palatable to all of them. Ideally, practice and evaluation would be merged into a single enterprise in which rigorous research becomes an integral component of program operation. It is unrealistic to think that most practioners will ever acquire all of the technical skills required to systematically conduct rigorous research on their activities (just as in is unrealistic to expect most researchers to become adroit practitioners). In addition, truely rigorous research is not always called for in the development of a project. When rigorous evaluation is called for, however, the PDE structure involving the collaboration of researchers and implementers appears helpful. Program Development Evaluation does not successfully resolve many of the sources of tension in merging action with evaluation, and it is a taxing procedure for the evaluator, but it is progress.

References

Blanton, J., & Alley S. Program development: A manual for organizational self-study. (Prepared under contract no. HSM-42-42-143 from NIMH). Nicasio, Calif: Social Action Research Center, 1975.

Chein, I., Cook, S., & Harding, J. The field of action research. American Psychologist. 1948, 3, 43-50.

Collier, J. United States Indian Administration as a laboratory of ethnic relations. Social Research, 1945, 12, 275-276.

Deese, J. Psychology as science and art. New York: Harcourt Brace Jovanovich, 1972.

Deutsch, M. Field theory in social psychology. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology (Vol. 1, 2nd ed.). Reading, Mass.: Addison-Wesley, 1968.

Dewey, J. How we think (rev. ed.). New York: Heath, 1933.

Empey, L. T. Field experimentation in criminal justice: Rationale and design. In M. W. Klein & K. S. Teilmann (Eds.), Handbook of criminal justice evaluation. Beverly Hills, Calif.: Sage, 1980.

French, W. L., & Bell, C. H., Jr. Organization development: Behavioral science interventions for organization improvement (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Glaser, D. The interplay of theory, issues, policy, and data. In M. W. Klein & K. S. Teilmann (Eds.), Handbook of criminal justice evaluation. Beverly Hills: Sage, 1980.

Hall, G. E., & Loucks, S. F. A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 1977, 14, 263-276.

Lewin, K. Action research and minority problems. Journal of Social Issues, 1946, 2, 34-46.

Lewin, K. Group decision and social change. In T. M. Newcomb & E. L. Hartley (Eds.), Readings in social psychology. New York: Holt, 1947.

Martin, S. E., Sechrest, L. B., & Redner, R. (Eds.). New directions in the rehabilitation of criminal offenders (Report of the Panel on Research on Rehabilitative Techniques). Washington, D.C.: National Academy Press, 1981.

Ogawa, D. K. Delinquency prevention: The record of accomplishment. This volume.

Office of Juvenile Justice and Delinquency Prevention. Program announcement: Prevention of delinquency through alternative education. Washington, D.C.: Author, 1980.

Pearl, A. Quality control in evaluative research of correctional programs. Paper presented at the National Institute on Crime and Delinquency, July 1962.

Perloff, R. (Ed.) Evaluator inventions: Pros and cons. Beverly Hills, Calif.: Sage, 1979. (Sage Research Progress Series in Evaluation, Vol. 2)

Quay, H. Three faces of evaluation: What can be expected to work. Criminal Justice and Behavior, 1977, 4, 341-354.

Sarason, S. B. The culture of the school and the problem of change. Boston: Allyn & Bacon, 1971.

Sechrest, L., White, S. O., & Brown, E. D. (Eds.). The rehabilitation of criminal offenders: Problems and prospects (Report of the Panel on Research on Rehabilitative Techniques). Washington, D.C.: National Academy of Sciences, 1979.

Tharp, R. G., & Gallimore, R. The ecology of program research and development: A model of evaluation succession. Unpublished manuscript. (Available from Roland Tharp, Department of Psychology, University of Hawaii, Honolulu 96822, or Roland Gallimore, Department of Psychiatry and Biobehavioral Sciences, UCLA, Westwood, California 90024.)

Wilkins, L. T., & Gottfredson, D. M. Research, demonstration, and social action. Davis, Calif.: National Council on Crime and Delinquency Research Center, 1969.

Footnotes

1. This use differs from some other common uses of the word objective--especially the usage of this term to imply a more specific restatement of a goal. A goal must itself be stated explicitly and measurably. Diffuse statements of general or difficult-to-specify aims might be called missions, but such vague aims are not to be confused with the crisp, clear statements sought when specifying goals and objectives.

2. Sometimes, a theory may imply a change in organizational structure as an objective, but objectives stated in structural terms should be carefully scrutinized to make sure they are not statements of interventions. If, for example, the government of a country were experiencing major civil unrest, it may assume that the problem is due to a lack of respect for government. It might then impose martial law in an attempt to restore civil order. Martial law, a structural change, might neither establish respect for government (an objective) nor civil order. If the establishment of martial law were viewed as an objective rather than as an intervention, useful information would be lost. Viewing martial law as an intervention is more useful. Similarly, the revolutionaries in this same country may want to overthrow the government (an intervention) presumably to achieve some objective, such as freedom, more equitable distribution of wealth, or a more satisfactory relation of workers to the mode of production. The distinction between a revolt and its objectives is an important one. Successfully implemented revolutions do not always increase freedom.

Accordingly the PDE structure makes a distinction between the objectives an intervention is intended to achieve and the intervention itself.

3. Forcefield analysis is useful to evaluators and project managers for an additional reason. Mapping the history of changes in the forcefield provides insight of potential general utility in planning future projects.

4. Strategies are used here as Lewin (1947) used the term "over-all plan." Such an overall plan is composed of steps

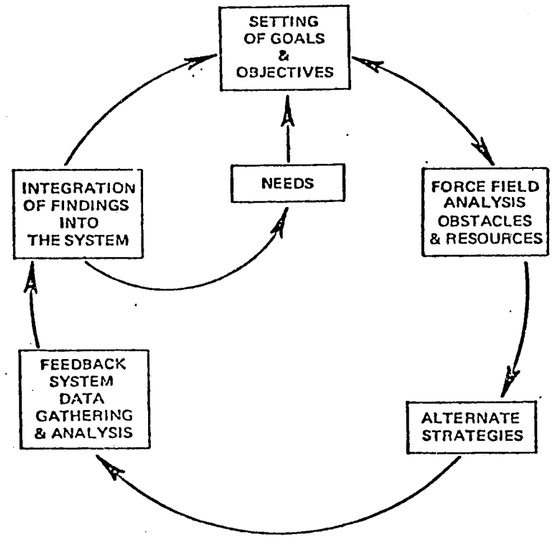
taken in succession, the consequences of each of which is in principle discoverable.

5. The concept of critical benchmark was incorporated in the PDE model at the suggestion of J. Douglas Grant. A critical benchmark is what Lewin (1947) called a "gate." "The constellation of the forces before and after the gate region are decisively different in such a way that the passing or not passing of a unit through the whole channel depends to a high degree on what happens in the gate region (p. 332)."

6. Tasks are related to what Lewin (1947) termed "action steps," but differ slightly as used here. The PDE model recognizes that a succession of tasks, any one of which may be fairly trivial in nature, are necessary to achieve some critical benchmarks. Often, however, strategies are sufficiently complex and the management of their execution sufficiently problematical that the specification of a number of tasks or "action steps" is a useful planning tool.

Figure 1

The Program Development Model



Source: Blanton & Alley (1975)

Figure 2

The Program Development Evaluation Model

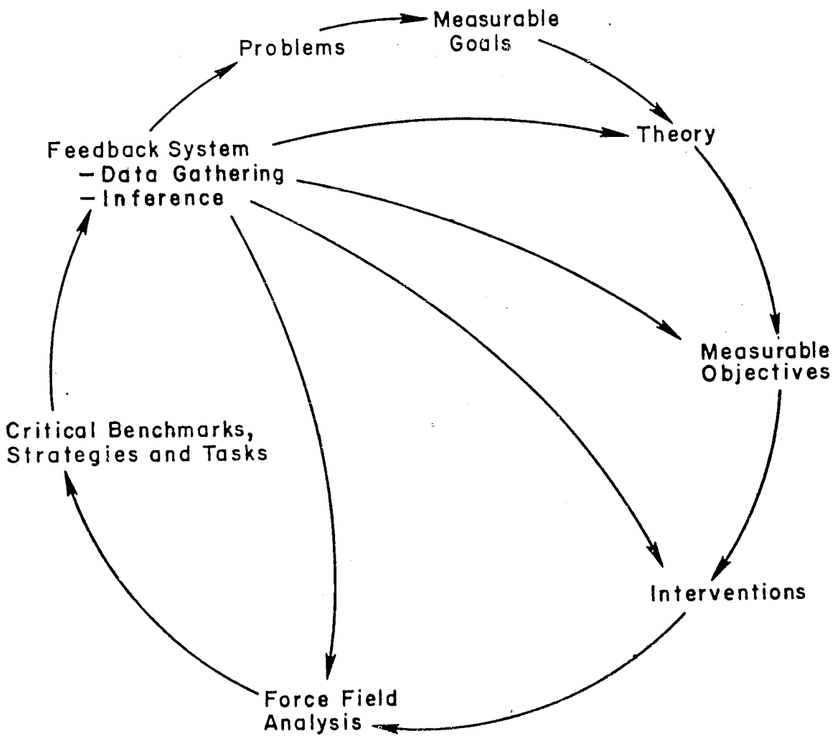


Table 1

PROGRAM DEVELOPMENT EVALUATION CONCEPTS

PROBLEMS: WHAT ARE THE MAJOR PROBLEMS, AND WHICH OF THESE ARE YOU ATTEMPTING TO ADDRESS?

GOALS: WHAT LONG-RANGE GOALS IS YOUR PROGRAM DESIGNED TO REACH?

(a) HOW CAN YOU MEASURE EACH GOAL? (HOW WILL YOU KNOW WHEN YOU HAVE REACHED EACH GOAL?

(b) WHEN DO YOU EXPECT TO HAVE MADE A SUBSTANTIAL DIFFERENCE?

(c) HOW WILL YOU KNOW YOUR PROGRAM MADE THE DIFFERENCE?

THEORY: WHY DO THE PROBLEMS EXIST?

OBJECTIVES: WHAT MEASURABLE CHANGES IN BEHAVIOR, ATTITUDE, OR ORGANIZATIONAL STRUCTURE DO YOU EXPECT TO BRING ABOUT?

(a) HOW CAN YOU MEASURE EACH OBJECTIVE?

(b) WHEN DO YOU EXPECT TO HAVE MADE A SUBSTANTIAL DIFFERENCE?

(c) HOW WILL YOU KNOW YOUR PROGRAM MADE THE DIFFERENCE?

INTERVENTIONS: WHAT ARE THE MAJOR PROGRAM COMPONENTS DESIGNED TO ACHIEVE THESE OBJECTIVES?

FORCE FIELD ANALYSIS: WHAT RESOURCES DO YOU HAVE AVAILABLE TO IMPLEMENT YOUR INTERVENTIONS?

WHAT OBSTACLES DO YOU ANTICIPATE ENCOUNTERING IN THE IMPLEMENTATION OF YOUR INTERVENTIONS?

CRITICAL BENCHMARKS: WHAT SPECIFIC MAJOR CHANGES MUST OCCUR TO IMPLEMENT YOUR INTERVENTIONS, AND WHEN MUST THEY OCCUR? (What needs to happen by when to get your interventions implemented?)

STRATEGIES: WHAT DEVELOPMENTAL STEPS WILL YOU TAKE?

TASKS: WHO WILL DO WHAT BY WHEN?

Students and Teachers in Context: The Measures used in the School Action Effectiveness Study

Gary D. Gottfredson, Deborah K. Ogawa, Donald E. Rickert, Jr., and Denise C. Gottfredson

Measurement is a central component of sound program development efforts, and measurement is essential in program evaluation. This report is a guide to using and interpreting measures of school climate and individual social development that may have broad applicability in school improvement efforts. It serves as a manual to help school administrators, counselors, psychologists, teachers, and school boards interpret survey information about school climates and about the characteristics of students and teachers.

Why Measure?

Every good administrator has some way of taking the pulse of his or her organization--of sensing when things are going well, when progress is being made, and of detecting problems or areas where change is needed. Measurement is a technique for making that kind of assessment. Classroom teachers routinely give tests to measure their students' progress in various curricular areas. This guides them in moving on to new curriculum, in re-emphasizing weak areas, and in meeting the educational needs of individual students. On a larger scale, economists routinely measure consumer prices and unemployment rates to help monitor the state of the economy.

In education, measurement has traditionally focused mostly on student ability, achievement, and interests. Educational decision-makers now have a large tool-kit of instruments to make measurements in these areas. These types of educational measurement have become so familiar to us that we now take them largely for granted. They help in making scores of decisions related to counseling, placement, and instruction, and are regular features in the landscape of educational evaluation.

These well-worn tools are limited, however, to a few areas. In other areas of concern to the education decision-maker, our measuring tools are less well

---

We are grateful for the comments of Michael Cook on a draft of this chapter. The development of this report was sponsored in part by Grant No. NIE-G-80-0113 from the National Institute of Education, U.S. Department of Education. The opinions expressed are solely the authors', and do not necessarily reflect the positions or policies of any agency.

CONTINUED

1 OF 3

developed. In some areas, such as psychosocial development, there are a variety of measuring devices available, but they are at present used primarily by the educational or psychological researcher. In this area we have, for example, Holland and Baird's (1968) Interpersonal Competency Scale, the Greenberger, Campbell, Sorensen, and O'Connor, J. (1971) Psychosocial Maturity Inventory, a variety of self-esteem scales, several measures of internal-external control, and other similar measures. All of these devices are potentially useful to the educational decision-maker in assessing interventions aimed at the specific characteristics of individuals that they purportedly measure.

Efforts have also been made to develop measures of school or classroom climate. Perhaps the best known examples of climate measures in the educational area are the commercially available scales produced by Moos and Trickett (1974). These scales are intended to measure structure, orderliness, and so forth. Unfortunately, they suffer from technical psychometric limitations (Richards, 1978), which result in needlessly confounding the measurement of environments with the measurement of individual differences within the environment. Furthermore, they are insufficiently comprehensive for the present purposes. But the measurement of school climates can now be improved by building on the research of Moos (1976), Stern (1970), Astin and Holland (1961), and Gottfredson and Daiger (1979), as well as the practical work of Fox and associates (1974).

A comprehensive and practically oriented set of climate measures can be of great value to educational decision-makers by making organizational diagnosis feasible. A comprehensive diagnosis of this type is useful in organizational self-study and for program planning because it can point out the strengths and weaknesses of a school's climate. And, a comprehensive climate assessment is of value in assessing progress toward improving the climate, in detecting changes in the climate due to "naturally occurring" events, and evaluating school improvement efforts.

Measuring Individuals and Organizations

This chapter describes a two-tiered set of measures devised to meet the needs of those educational decision-makers who seek to improve education for individual students, or to improve school climate more generally. One tier assesses the characteristics of individual students and individual teachers that are relevant to organizational climate, or to important personal outcomes. The other consists of school-level climate measures that directly assess some important dimensions in which schools vary.

The measures are divided into these two classifications for an important reason. We have all experienced differences in the psychosocial climates of different organizations, and we can easily appreciate that organizations differ in the environments that they provide. Yet we also know that different individuals often have different views of the characteristics of the same organization. Therefore, in assessing a given climate, it is important to average across many different reports-in essence treating individual differences as error. These differences are, however, the very reason we measure individuals. Accordingly, two distinct sets of measures are called for. Besides the generalized assessments, individual measures are needed for personalizing instruction and for comparing the effectiveness of alternative educational treatments.

The measures described here were developed specifically for the School Action Effectiveness Study (SAES) because no comprehensive and psychometrically adequate battery was available elsewhere. They are rooted directly in a program of research on delinquency and school environments conducted over the past several years at the Center for Social Organization of Schools. The instruments used are based in part on those used in the National Institute of Education's (1978) Safe School Study, instruments suggested by Fox and associates (1974), the School Initiative Evaluation questionnaires, and a number of other instruments used in major social surveys in recent years. In addition, relevant items (with necessary modifications) from other devices are used. Decisions about useful measures are based on a review of the goals and objectives of the particular alternative education project being evaluated, on Gottfredson's (1981b) account of delinquency theory and strategies for organizational change, and on many discussions with action project personnel--using the Program Development Evaluation framework--of the goals and objectives of their particular school change efforts.

Some Essential Psychometric Concepts

In order to use the measures about to be described in an informed manner, it is important to understand several ideas: (a) the relative nature of psychosocial measurement, (b) reliability and (c) construct validity.<1> The following paragraphs review these ideas.

Relative Measurement

We have relatively few absolute measures in social science. In other words, simple counts of "units" of achievement or interpersonal competency or fairness are impossible to obtain. Instead, we typically express their levels in relative terms. For exam-

ple, achievement test results are often presented in terms of percentile rank or standard score form. These forms of expression involve statements of the standing of an individual (or organization) relative to some norm group of people (or organizations). For example, a percentile rank of 76 on an individual test would mean that out of 100 individuals representative of the population on which the test's norms are based, 76 persons would have a score lower than this one.

Sometimes, too, some form of "scaled score" or "standard score" is used. Often, raw scores are converted to rescaled scores with a mean of 50 and a standard deviation of 10. (The mean is the arithmetic average of a set of scores, and a standard deviation is a unit of dispersion or spread.) This way of expressing scores is especially useful when it turns out that scores have the familiar bell-shaped distribution; that is, when lots of people (or organizations) have scores near the middle of the distribution, and the relative frequency of scores trails off symmetrically for higher and lower scores. (Roughly, this is what is meant by a "normal" distribution.) This kind of standard score is easy to interpret, because about 68% of all persons (or organizations) in the sample used to construct the norms have scores between 40 and 60. Figure 1 helps show how easily standard scores of this kind can be interpreted. About 95% of all scores fall between 30 and 70. Scores below 30 are relatively rare, and scores above 70 are also relatively rare. Comparatively, such scores are exceptionally high or exceptionally low. In this report we will use both percentile ranks and standard scores to present results. These simple forms of presentation were chosen in part because they are familiar to anyone who has interpreted the Scholastic Aptitude Test, which reports standard scores with a mean of 500 and a standard deviation of 100, and which also reports percentile ranks.

In interpreting such scores it is important to bear in mind that they express scores relative to the norm group. Different norm groups will themselves differ somewhat in their means scores (and also in their dispersion). Therefore a score that is, for example, at the 65th percentile relative to one norm group could be at the 30th percentile relative to another norm group. There is no such thing as a magically "correct" or even "most appropriate" norm group. In this report, the norm group is the total population of students, teachers, or schools involved in SAES, unless otherwise explicitly stated. Therefore, a "high" score is a high score relative to these groups.

Not all of the attributes of individuals and organizations that are important to

measure are normally distributed. Sometimes scores tend to pile up at the top or bottom of a scale, and gradually trail off towards the other end. Roughly speaking, this is what is meant by a skewed distribution. In such a distribution, interpretations based on the assumption of a normal distribution can be somewhat misleading. Therefore, we sometimes point out the skewness of a distribution to aid in interpretation. For example, delinquent behavior tends to show a markedly skewed distribution, with many people reporting few delinquent acts (or earning "low" scores), and a very few individuals reporting a great many delinquent acts. Scores pile up at the bottom of the scale.

The schools, and hence the students and teachers, involved in this program may be expected to differ from nationally representative samplings. In particular, one of the selection criterion for participation in this program was a credible indication that the problems of crime, dropping out, or nonattendance were great in these schools or their cities. In addition, inner-city, predominantly minority schools are clearly overrepresented. We know from other research (e.g., Gottfredson & Daiger, 1979; Gottfredson, Joffe, & Gottfredson, 1981) that such schools on average experience more disruption than other public schools in the nation.

To provide some perspective on the way this norm group may differ from a nationally representative sample of schools, we have prepared Table 1. It shows selected characteristics of the SAES sample and nearly nationally representative samples of youth. The SAES schools enroll much larger Spanish-American and Black populations than the proportion of these ethnic groups in the nation, and a much lower percentage of students enrolled in SAES schools live in intact families. Although victimized themselves to roughly the same extent as typical students, these students engage in much more delinquent activity than do typical youth.

Please note that the psychometric use of the word "norms" has little or nothing to do with some everyday language uses of the word. In everyday language we sometimes use "norm" to mean an ideal or required standard. So Derek may be well below the weight norms for women of her height, but there appears to be general agreement that her physique is not otherwise "substandard." Similarly, it is quite possible for a school to have students who show an "average" degree of satisfaction with school but who are rather uncomfortable--or who are average in reading achievement according to large city norms, but who do not read well at all. In interpreting any particular results, readers should probably consider both their own "ideal" norms and the

"statistical" norms presented here.

Reliability

Chance, sloppiness, ambiguity, temporal instability, and heterogeneity of meaning or interpretation can influence any measure. Measurements of the distance between Baltimore and New York made by the odometers in a number of different cars would tend to agree pretty well, but not perfectly. They would have high, but not perfect, reliability. Reliability is a technical term used to describe the relative contributions of measurement error and "true" score variability to a scale or other measure. Technically, reliability is the proportion of the variance (a statistic summarizing variability) in a score that is not error. Because there are many ways of defining error, there are many ways of estimating reliability (Stanley, 1971). The reliability coefficients reported in this manual (alpha) are based on the analysis of items administered on a single occasion and therefore exclude temporal instability from the definition of error. They can be interpreted as an index of how well the scales measure whatever they measure at a given point in time.

Knowledge of the reliability of a test or other index is important because low reliability means that the device does not measure anything well. A high reliability means that the device measures something. (What that something is, is what construct validity is all about.) Reliability coefficients can range from 0 to 1.0. A reliability of 1.0 is high, meaning that the score contains no error. Over the years practitioners have developed rules of thumb for acceptable levels of reliability for different purposes. In general, it is not sound practice to use tests with reliabilities much below .7 or .8 for individual diagnosis, personnel decisions, and so forth. When interpretations of patterns or profiles are to be made, it is especially important that reliability be this high, or higher.

For evaluation purposes, lower levels of reliability of measurement at the individual level are acceaptable and are sometimes to be preferred, because of three related considerations. First, because the scores of many individuals are usually averaged in an evaluation, dependable estimates of true-score means can be obtained even with rather unreliable individual measures (see Stanley, 1971). Second, the longer the scale (i.e., the more items), the more reliable it is, other things being equal, but it is often difficult, time consuming, or costly to administer long scales. As an alternative, using short scales with many persons gains good estimates of group means. Third, in an evaluation it is necessary to measure many things.

This is because action programs have many goals and objectives, and because it is always wise to search for unanticipated positive outcomes or side-effects of a program. But administering many highly reliable (i.e., long) scales is prohibitive. Fortunately, a large test group again comes to the rescue. Using short scales with many people solves the problem and yields satisfactory estimates of true-score means.

As a rule of thumb, scales with reliabilities as low as .5 are adequate for use in an evaluation, provided that the project being evaluated uses randomization as a selection device, or that any selection is absolutely independent of (i.e., unrelated to) the goals or objectives of the program. In such an evaluation, it is not necessary to attempt to adjust for pre-existing or spurious group differences on outcomes. When it is necessary to make such adjustments by using statistical "controls," reliabilities for the control variables must be as high as possible. The rule of .5 is too lax in this case because when the "control" variables are unreliable they do an inadequate job of correcting for spurious differences between groups. Therefore, to enable a sound evaluation, a project which does not randomize should use more reliable (i.e., longer) scales encompassing measures of all relevant characteristics in which the treatment and comparison groups may differ.

Validity

Validity has to do with the meaning and interpretation of an index or score, and is closely linked with theory. Theory involves constructs or ideas about the causes or nature of phenomena. Often, measurement has meaning only in the context of some theory. For example, some educators have a theory that a general ability called intelligence underlies much human performance, or a least scholastic achievement. The measurement of intelligence using a paper and pencil verbal ability test may make sense in terms of this theory. Because the theory predicts that this test will correlate with school grades, evidence about the validity of a test for measuring the construct of intelligence can come from an examination of the empirical relation between test scores and school grades. The same evidence provides information about the utility of the theory. Theories and tests are thus validated in a common process. We speak of a test as validated when empirical evidence has in general shown the test results to follow the predictions of a theory that has been useful.

In addition, when there is agreement about what a construct means, some evidence about validity can come from an examination of the item content of a test. For example, most of us would probably agree that a test to see how many bricks

a person can load on a truck in an hour is a poor test of verbal ability, and that a list of multiple-choice vocabulary items would provide a more valid measure of that construct. (Similarly, the vocabulary test would be a poor test of endurance.) Therefore, deliberately including items to measure a given construct in itself can provide some limited degree of confidence in a scale's construct validity.

The evidence is strengthened if the scale shows expected patterns of correlations with other scales. And it is especially strengthened if applicable experimental manipulations influence scores in predicted ways. Other evidence of validity can come from an examination of differences in scores on the scale among groups known or believed to differ in the characteristic being measured. For widely used instruments, these kinds of evidence accumulate over time. Eventually, a basis for judgment about a scale's construct validity emerges--although different judges often disagree.

There can therefore be no such thing as an absolutely validated test or scale.

Subsequent sections describe the origins, development, and psychometric properties of a multi-level set of assessments of schools and their inhabitants. These sections are intended to provide information about reliability and validity, and to describe the normative interpretation of these assessments.

Measures of Students

Five sets of measures of individual students have been developed to measure (a) constructs suggested by the staff of action projects, and (b) delinquency prevention theory. These measures of students are needed to assess project effectiveness under difficult field research conditions and to learn more about what works for whom.

Social Background

Measures of social background or family characteristics are needed for two reasons: (a) They provide essential statistical controls to aid in demonstrating project effectiveness when evaluation designs calling for statistical adjustments are necessary, or when stronger designs fall apart. (b) In a few cases, projects aim to alter family characteristics--usually the extent to which parents value education or encourage their children to perform well in school.

Accordingly, the following six measures were developed:

Parental Education. This two-item scale is based on decades of research that show parental education to be a powerful antecedent of schooling outcomes, especially of persistence in education (Sewell, Haller, & Portes, 1969a). The two

items ask how much education a student's father and mother completed. The scale has a reliability coefficient of .76 and, with the exception of the two ethnic groups for whom only very small samples are available, has approximately equal reliability for all race-sex subgroups examined. Table 2 displays the scale's reliability estimates for ten groups. (At a later time, parental occupational level will be added to this scale, as much research (Duncan, 1961; Treiman, 1977) indicates is appropriate. Write-in data about parental occupation require much time to process.) This measure may be taken as an indicator of family socio-economic status. It is known to be a good predictor of schooling outcomes such as persistence and grades (Bachman, O'Malley & Johnston, 1978; Jencks, 1979), but it is usually only weakly related to delinquent behavior at the individual level--although perhaps it has a stronger relation to more serious delinquency (Tittle & Villemez, 1978; Gottfredson, 1981a).

Parental Emphasis on Education. This four-item scale asks for information about the degree of parental attention to the student's school performance and parental expectations for school persistence. It was suggested by action project theories that attributed student non-attendance to a lack of parental encouragement or "value" on education. And, parental influence is demonstrably predictive of student persistence in school (Otto, 1976). The scale is only moderately reliable--.46 overall, with coefficients ranging from .38 to .62 for race-sex subgroups. Reliability coefficients are only slightly lower for black subsamples, where single-parent (female-headed) family structures make the potential differential utility of this scale a matter for concern. Modest differences among groups exist, with American Indian males reporting rather low parental emphasis on education relative to other groups. (See Table 3). The scale has moderate negative correlations with self-reported delinquency (see Table 4), and has an expected, but small, positive correlation with student reports of effort spent on school work.

Maternal Role Model (Negative). Although a sensitive matter to address in a survey used in schools--especially where administrators are uncertain of the confidence of their communities--this scale is intended to tap an important theoretical antecedent of delinquent behavior. Social learning theory (Bandura, 1971) and differential association theory (Sutherland & Cressey, 1955) appear to imply that a person learns to behave in accord with models in the person's environment. This five-item scale contains items asking whether the person's mother gets mad a lot, drinks too much, or spends all her money on herself. It is

characterized by low item responses (i.e., most students say "no" to most items), and therefore has a low reliability--.36 overall, with male white and American Indian reliabilities very low. Despite its low reliability, it correlates .22 with self-reported delinquent behavior, making it a potentially useful statistical control variable in weak evaluation designs.

Maternal Role Model (Positive). This scale is composed of items describing positive aspects of the maternal role model. These items were included largely to soften the impact of the items in the negative role-model scale just described. The scale has a modest to dismal degree of reliability (especially for white and Indian males), and this set of items (mother is a hard worker, fixes things around the house, etc.) is relatively uncorrelated with the Negative Maternal Role Model Scale.

Paternal Role Model (Negative). This scale parallels the corresponding Maternal Role Model Scale. It has somewhat higher reliabilities than that scale, and is moderately correlated with self-reported delinquent behavior.

Paternal Role Model (Positive). This scale parallels the corresponding maternal scale in intent and in psychometric characteristics. It is only modestly correlated with the corresponding maternal scale.

Table 5 shows correlations among the family background scales. These correlations are modest, and low relative to the scales' reliabilities--implying that the scales each measure a relatively independent dimension of family social background.

Social Relations

Three measures of a student's social relations were developed because of (a) empirical and theoretical links between bonds of affection or respect for others and conforming (non-delinquent) behavior, (b) powerful statistical associations between delinquent behavior and delinquent peer influence, (c) the central place given to peer influence in the theories of several of the action projects, and (d) the explicit assumption made by several projects that parental supervision governs student attendance. One of these scales (attachment to parents) is intended for use as a statistical control variable to strengthen evaluation designs. The others measure outcomes of importance for all or some projects.

Attachment to Parents. This scale, intended to measure Hirschi's construct of the same name, incorporates several items closely related to items shown in earlier studies to be correlated with delinquent behavior (Hirschi, 1969; Hindelang, Hirschi, & Weis, 1981; D. Gottfredson, 1981b). An

attempt has been made to engineer a more potent scale by including more items related to this construct. The six-item scale, asking students how close they are to their parents, how much they like them, and so forth, has an overall reliability of .61. It correlates as expected with self-reported delinquent behavior (see Table 4), in accord with Hirschi's (1969) theory that attachment to parents creates a stake in conforming behavior. This agreement provides some evidence of the construct validity of this scale.

Negative Peer Influence. This scale measures a construct central to the explanations of delinquency and non-attendance formulated by several of the action projects. It is rooted directly in earlier research (summarized by Empey, 1978) that shows delinquent peer associations to be powerful predictors of delinquent involvement. In addition, it incorporates items related to dropout, similar to those used in earlier studies of persistence in schooling (Bachman et al., 1978). It is, however, an attempt to engineer a longer, more powerful, and broader-based measure of negative peer influence. This nine-item scale has reliabilties ranging from .53 to .74 across subgroups (Table 2) and, as expected, is a potent correlate of delinquent behavior (Table 4). It contains items asking whether the student's best friend is interested in school, thinks getting good grades is important, thinks school is a pain, or has been involved in delinquent activities.

Parental Supervision. Although we know of no clear evidence that parental supervision is related to delinquency, at least two of the action projects attribute truancy and nonattendance in part to a lack of parental supervision. Therefore, we attempted to build a scale tapping student reports on whether their parents usually know where they are and what they are doing, and whether as far as their parents are concerned they are free to come and go as they choose. This scale, with only two items, has low reliability. It does, however, correlate with delinquent behavior and with students' efforts in school, as predicted by the action project theories (Table 4).

The correlations among these measures of social relations are presented in Table 6. It shows that parental supervision, which forms only a weak scale, is correlated as expected with the other measures, but that these correlations are close to its reliability, raising doubts about the usefulness of maintaining this scale as a separate entity. Its correlation with delinquent behavior is so high, however, that the measure is of some value as a statistical control. This scale should be improved if possible.

Attitudes and Social Development

Social development is a major goal of the Alternative Education Program. In this area, there was considerable prior work to build on in choosing measures to include in the battery.

Alienation. The four-item Alienation Scale is based in part on Srole's (1956) Anomia Scale, but fewer items are included, and the wording of itmes has been changed to give them more school-related content and to make them sound a little less bizarre. Specifically, alienation items used in the School Initiative Evaluation (Grant et al., 1979) were modified for use here. Items include, "These days I get the feeling that I'm just not a part of things." And, "I feel no one really cares much about what happens to me." Overall, this short scale has a reliability of .44, and works about equally well for all groups examined except for the very small sample of Asian males, who score quite low on the average. As expected, the scale correlates positively with self-reported delinquent behavior, and negatively with reports of effort expended on school work (see Table 4).

Attachment to School. This is a central construct for many projects whose major goal or objective is the development of positive student attitudes toward school. The construct is also central to social control theories of delinquency (Hirschi, 1969) that view attachment to school as a major social bond restraining individuals from participation in delinquent behavior. Consequently, we have constructed a relatively long and broad-based measure of attachment to school. This 10-item scale has reliabilities ranging from .67 to .81 across subgroups--.75 overall. Items ask the students if they like the school, if they like the classes, how important getting good grades are, and so forth. The scale is, as expected, a powerful correlate of delinquent behavior (negative) and effort expended at schoolwork (positive). Other correlations (not shown in a table) indicate that attachment to school is also related to school attendance.

Belief. The expectation that individuals differ in the extent to which they believe in the moral validity of conventional social rules, and that the degree of belief influences behavior, is widely shared. A common goal of peer-group-based interventions to prevent delinquency is to strengthen belief by using peer pressure. The item content of Gough's (1964) socialization scale--which was developed through empirical efforts to discriminate between offenders and non-offenders--lends support to this popular notion. And, belief is a central construct in social control

theory, which postulates that people differ in the degree to which they have internalized rules, and that they therefore are constrained from involvement in delinquent behavior to different degrees. Much empirical evidence supports this idea (e.g., Wiatrowski & Swatko, 1980; D. Gottfredson, 1981b; Hirschi, 1969).

Consequently, in order to measure this aspect of social development we have assembled a short scale from well-worn items used in other surveys, whose characteristics were known. The six-item scale contains items such as, "It is all right to get around the law if you can;" "Taking things from stores doesn't hurt anyone;" and "People who leave things around deserve it if their things get taken." The scale has a reliability of about .50, and it correlates .27 with self-reported delinquent behavior, as earlier research and theory predict it should.

Interpersonal Competency. This scale is composed of four items from Holland and Baird's (1968) Interpersonal Competency Scale. That scale has well-studied psychometric properties. It consistently has moderate reliability and correlates positively with other measures of psychological health or adjustment, and negatively with measures of alienation. The fifth item was written by Holland especially for the present purpose, to give the scale more school-related content. It has a reliability coefficient overall of .42. This measure of social development correlates positively with reported effort expended on school work, and it is nearly independent (uncorrelated with) self-reported delinquent behavior. This accords with other evidence that delinquent involvement is only modestly associated with psychological health (Waldo & Dinitz, 1967), although there are some alternative views (e.g., Quay, 1964).

Involvement. This scale is intended to measure a central construct in social control theory that does not appear to have been well measured in the past. The idea is that involvement in conventional activities creates a stake in conformity, because a person involved in rewarding activities has something to lose by misconduct. This scale (not to be confused with environmental measures of student influence or involvement in decision-making) is composed of 15 items (most of which were adapted from the current National Longitudinal Study questionnaire) asking about a student's participation in a wide variety of in-school and out-of-school activities, including school-sponsored activities and work. It has an overall reliability of .75, but does not correlate as expected with reports of delinquent behavior, casting some doubt on its construct validity. It does, however, correlate .31 with students' reports of rewards they have received

in school, and has a small .12 correlation with a measure of practical knowledge (to be described shortly). Although this scale was intended to serve as an important intermediary outcome measure, its utility is in doubt.

Positive Self-Concept. A number of self-esteem scales with well-researched properties (Robinson & Shaver, 1973, review more than 30 measures) are available. To create a short scale, items previously used by Rosenberg (1965) and an item similar to one used by Coopersmith (1967) were subjected to analysis along with another set of items constructed to capture aspects of self-concept specific to schooling and delinquency. This scale has its base in the labeling perspective (Lemmert, 1972), which implies that if people are treated as slow learners or delinquents, they will come to incorporate aspects of those social definitions into their own self-concepts. Positive self-concept, therefore, is an important intermediary outcome according to this perspective. Effective alternative education projects would presumably increase scores on the positive self-concept scale, and a program with unexpected negative side-effects could decrease scores. Item analysis did not justify treating self-esteem as a separate scale from these labeling outcomes, because items are about equally correlated across the two sets. Weak items were excluded, leaving a 12-item scale with reliabilities ranging from .51 to .78 across subgroups, .63 overall. (In these as in other item analyses, statistics for all subgroups were examined to ensure that items worked generally across groups.) Items include, "My teachers think I am a slow learner;" "Sometimes I think I am no good at all;" "I am the kind of person who will always be able to make it if I try;" and "I do not mind stealing from someone--that is just the kind of person I am." The scale correlates .48 with reported effort on school work, and -.24 with self-reported delinquency. It also correlates -.39 with alienation and .39 with interpersonal competency, lending support to its construct validity.

Practical Knowledge. To provide a simple measure of self-reported competencies needed for coping with everyday life, a seven-item measure was created for the survey. Although this self-report scale may be a poor substitute for a more comprehensive or task sample approach, it seemed the only way to build a measure of this kind of social development into a multi-purpose battery. The scale works remarkably well, with a reliability coefficient of .71, and good item properties across all groups studied. It is relatively independent of the other measures of attitudes and social development, and of self-reported delinquent behavior. Because it has

not been well-studied, it should be interpreted cautiously. We will be able to report more information about this scale after other data are merged with the questionnaire measures.

Rebellious Autonomy. In talking with persons running action projects, especially the Peer Culture Development Project in Chicago, explanations of the problem of delinquency sometimes involved a kind of peer or gang culture that resembles Miller's (1958) characterization of subcultural socialization. The peer or gang culture may incorporate a set of socially-shared expectations that are different from what might be called middle-class expectations. Differences may be so great that in behaving according to the "lower-class" system a person may violate norms of middle-class culture, and may appear to be deliberately non-conforming or malicious to a "middle-class" observer. In particular, middle class concerns with achievement may not be shared by "lower class" youth (cf. Attachment to School and Educational Expectations). Instead, these "lower-class" youths, according to Miller are concerned with trouble, toughness, smartness (i.e., manipulative skill), excitement, fate (explaining events by reference to chance or luck), and autonomy (an ambivalent relation to authority--overtly desiring not to be pushed around but covertly desiring to be cared for and controlled).

Because of this recurrent theme in our discussions with action project personnel, it seemed important to incorporate brief measures of this type of "subcultural" value system. Unfortunately, we know of no devices already in existence designed to measure this constellation. But we could locate isolated items from previous surveys whose distributional properties could be examined, and we found some interview quotes that suggested some items. We then wrote, therefore, nine items that seemed to capture the essence of these themes. Item analyses implied that three of these items formed a scale for all race-sex subgroups. The deletion of poor items, however, narrowed the content of the set down to items that appear to reflect a rebellious autonomy: "Whether or not I spend time on homework is my own business;" "I should not have to explain to anyone how I spend my money;" and "I don't like anybody telling me what to do." The scale has a reliability of .46 overall. The scale correlates .18 with self-reported delinquent behavior and -.24 with belief.

The correlations among the eight measures of attitudes and social development are shown in Table 7. These correlations indicate that Alienation is closely (negatively) related to Attachment to School. Indeed, the correlation between these two scales is higher than the reliability of the former scale. This implies

that some parsimony would result from combining these scales. We report them separately only because we believe users will like to see them separately. We do not recommend making substantive interpretations of differences between these two scales. The other scales are all relatively independent, implying that they measure fairly distinct aspects of attitudes and social development.

Self-Reported Behavior

School Effort. That students who earn low grades in school tend to drop out of school and to engage in delinquent behavior more than others are two of the best documented and consistent findings in the literature (Gottfredson, 1981a). Social class and ability are modestly associated with these same outcomes but do not completely account for these associations. Therefore, it seems likely that these outcomes are determined at least in part by grades--the major, if infrequently applied, reward system of traditional schooling. Grades in school are not determined solely by ability and social class, of course. Industrial psychology's expectancy theory (Porter & Lawler, 1968) suggests a mechanism whereby effort is expended if valued rewards are perceived as attainable, and in which effort is one of the determinants of both performance and rewards. Therefore effort is an important intermediary outcome variable that should be assessed in the evaluation of a program designed to prevent delinquency and foster persistence in schooling.

Unable to locate existing questionnaire measures of this construct, we developed one. This five-item scale has a reliability of .51 overall. (It is somewhat less reliable for Spanish Americans in this sample, presumably because some of its item content deals with homework, which is rarely assigned in Puerto Rico where most of the Spanish Americans in the sample attend school.) The scale includes these items: "Compared to other students, how hard do you work in school?"; "I turn my homework in on time"; and "I don't bother with homework or class assignments." As expected, females score higher on this scale on the average than do males. It correlates .39 with self-reported grades and .34 with attachment to school, supporting its interpretation as a measure of effort expended on school work.

School Non-Attendance. The Alternative Education Program is intended to demonstrate and evaluate projects that aim to increase attendance. Dependable attendance data are not always available from school records, so we decided to incorporate a brief self-report measure of attendance in the questionnaire to provide back-up data. This decision proved to be wise: we were unable to obtain these data from records for

at least one project, the data from many others is so erratic or incomplete as to be of limited value, and some attendance data that will ultimately prove valuable will require extensive editing before it is usable.

Two items, one asking how often the student cuts school all day and one asking about class skipping, compose this brief scale, with an overall reliability of .66.

Self-Reported Delinquency (Total). One way to find out what people do is to ask them. Naturally, not everyone tells the truth, perhaps especially when the questions are sensitive. A common assumption is that people will conceal information about their participation in illegal behavior, and so under-report. At the same time, the rates of delinquent behavior estimated by the self-report method are higher than those derived from official records (Empey & Erikson, 1966). There is thus a great deal of debate among criminologists about the appropriate way to measure criminal behavior.

Although there is no need to go into the arguments in any detail here, a major issue is that typical self-report measures (e.g., Nye, 1958) tend to measure minor "offenses," some of which are not "crimes," or would not be crimes if committed by an adult. Elliott and Ageton (1980) have recently presented evidence that self-report scales involving more serious offenses tend to resemble measures based on official data more than do scales involving only trivial items. Hindelang, Hirschi, and Weis (1981) have also just published the results of a major examination of self-report delinquency measurement.

The bottom line, insofar as it can be perceived at present, is that fairly long, variety-type scales involving a range of serious delinquent behavior do produce results that parallel official records for some subgroups but not for others. Hindelang et al. (1981) report validity coefficients for a number of variables that imply very low validity of self-reported data for officially "delinquent" black males, and much better validity for other subgroups. This is a difficulty that should be kept in mind in interpreting these self-reported data. It appears related to a similar problem of differential reliability in studies of educational persistence (Bielby, Hauser, & Featherman, 1977; D. Gottfredson, 1981a), and it points out the importance of obtaining official data for purposes of evaluation.

The specific self-report measures used here are modified from those used by Elliott and Ageton (1980) and by Hindelang, Hirschi, and Weis (1981). Many of Elliott's items were used, but pretests showed that the response formats created problems in group question-

naire administration. At about the same time, a manuscript of the Hindelang et al. book became available, in which the authors observed that "ever variety" and "last-year variety" scales yielded very good results (and a consultant's--La Mar Empey's--advice suggested the same). For purposes of evaluation, "ever variety" items (which measure prevalence rather than incidence) could not be used. Therefore we designed "last-year variety" items. These items ask, "In the last year have you..." Respondents indicate, for example, whether they have "stolen or tried to steal something worth more than $50."

A 19-item scale constructed in this way has very nice characteristics--considering that only a small proportion of respondents answer yes to any given question. Overall, reliability is .84. In the hold-out sample--those not used to conduct item analyses, but set aside to obtain unbiased estimates in a new sample--the subgroup reliabilities range from .63 to .93. The single low coefficient is for Asian-American females, who report almost no delinquent behavior. These reliabilities compare favorably to those obtained by Hindelang et al. (1981) with a 63-item last-year variety scale--.83 to .92 for black and white males and females.

Readers interested in a more thorough understanding of scales of this type and their relation to other variables of interest should see Hindelang et al. (1981).

Self-Reported Drug Use. Action project personnel have shown considerable interest in a component of delinquent behavior involving drug use. To provide a measure to meet their needs, we have also scored a five-item subset of the longer (total) S-R delinquency scale. It is composed of items asking about the use of cigarettes, liquor, marijuana, and other drugs, and about going to school "high." (A sixth item about glue sniffing was left out because the analyses did not support its inclusion for all ethnic groups.) This group of items closely resembles the Hindelang et al. (1981) Drug Index. It has an overall reliability of .84.

Self-Reported Serious Delinquency. A second subscale was constructed to measure only conduct that nearly everyone would regard as criminal. It includes 11 items (including one about selling drugs that Hindelang et al. would place in the drug cluster) and has an overall reliability of .79.

Measures of School Experiences

It is anticipated that the projects in the Alternative Education Program will expand the range of school rewards beyond those represented by traditional classroom grades. Accordingly, we have developed two scales to measure students'

rewarding and punishing experiences, in an effort to assess this important but hard-to-measure set of outcomes. School rewards and punishments make sense intuitively as probable causes of school attachment, effort, and persistence.

School Punishments. This four-item scale is an index of the negative sanctions an individual student experiences. It asks whether the student was required to stay after school, given an extra assignment, or had his or her grade lowered as a punishment. Its reliability coefficient for the total sample is .54; according to this index males experience more punishment, as expected. The scale correlates .30 with self-reported delinquency, -.28 with positive self-concept, -.30 with belief, -.22 with school effort, and .24 with negative peer influence.

School Rewards. This six-item scale is an index of the positive sanctions an individual student experiences. It includes reports of incidents in which the teacher complimented the student's work, the student was given a prize or award, or the student won an award for his or her class. The reliability coefficient for the entire sample is .54. The scale is relatively independent of sex, and is correlated .25 with school attachment.

Victimization. A final measure of school experiences deals with personal victimization. It is intended for use in assessing the amount of crime in the environment, and it is used in the aggregate to characterize the school. Nevertheless, the scale's characteristics at the individual level are of some interest. Containing five items, the scale has a reliability coefficient of .47. As earlier research (National Institute of Education, 1978) showed, boys are victimized more often than are girls. Victimization is correlated .24 with self-reported delinquency, implying a moderate tendency for persons who are victimized to engage more readily in delinquent behavior themselves. It correlates -.27 with school attachment and -.28 with self-esteem. Interestingly, its highest correlate among the variables examined is punishment (.35): Students who report more frequent personal victimization also more often report being punished in school.

Invalidity. There is always some concern that students may not faithfully complete their questionnaires, that they may fool around or give silly answers. As a check on this, a scale was included to detect unusual or nonsensical responses. This five-item scale is composed of items that a careful respondent would answer in only one way. It is keyed so that a rare response earns a point. This scale has not yet been used, but will be used later as a check on the results and as a quality control mechanism.

The item content of all student-level scales is shown in Appendix Table A.

Measures of Teachers

The second largest group of inhabitants of a school environment are the teachers who work there. Students in the aggregate help to create an environment for the teachers, just as teachers create an environment for the students. A characterization of the teachers is important in describing a school or a project.

Several of the action projects' theories lead to interventions geared toward teachers. The interventions are intended to improve classroom management, to change teachers' attitudes, or to involve them in new kinds of activities. One aspect of the evaluation therefore involves the measurement of teacher characteristics. Before turning to our account of school environmental measures, we will describe a set of individual-level teacher measures.

Pro-integration Attitude. This four-item scale is a measure of attitudes toward integrated education. It is included because these delinquency and school improvement programs are designed to provide services to heterogeneous groups of students. One component of several projects is training teachers to manage heterogeneous classrooms and to interact with a variety of kinds of students. This scale is expected to be useful in assessing the effectiveness of teacher participation in such activities. It has a reliability coefficient of .69 (Table 9) and is relatively independent of the other teacher scale (Tables 10). As might be expected, nonwhites tend to score somewhat higher than whites on this scale.

Job Satisfaction. This scale is composed of three of the four items in Hoppock's (1935) scale of the same name, which has been used widely in research and has demonstrated impressive evidence of convergent validity (Robinson, Athanasiou, & Head, 1969). Even shortened to three items it has a reliability of .80. It may confidently be taken as a measure of how well teachers like their jobs.

Interaction with Students. This six-item scale measures the extent of out-of-class interaction that a teacher has with students. Items ask about tutoring individual students before or after school and discussing their personal problems with them. It has a reliability coefficient of .67, and correlates positively with Job Satisfaction, negatively with reports of classroom disruption, and positively with the extent of recent continuing education activities.

Type A Sanctions. This is one of two scales developed in an attempt to describe the types of responses to student conduct used by the classroom teacher. We are unaware of

any short questionnaire measures of this aspect of classroom management, but provocative evidence from earlier research (McPartland & McDill, 1977; Gottfredson & Daiger, 1979) suggests that responses to conduct are important in preventing disruption. Therefore we used the best advice we could get to develop lists of various ways classroom teachers might respond to student behavior. These lists became items in the questionnaire. Through factor analytic examination and internal consistency item analysis, two scales emerged.

The first set of items is termed "Type A" Sanctions. A teacher who reports lowering grades as a punishment, sending misbehaving students out of class, and paddling or reprimanding the students in class is given a high score. The scale has a reliability of .47. Its largest correlate among the other teacher measures is the amount of disruption the teacher reports; it is also moderately negatively correlated with nonauthoritarian attitudes.

Type B Sanctions. This scale was developed in the same way. In contrast to the Type A scale, which seems to include responses rooted in frustration, Type B Sanctions appear to involve a wider range of resources. To earn a high score on this scale, a teacher reports giving extra schoolwork, awarding special privileges for good behavior, taking away privileges for misconduct, calling parents, and referring students to the counselor or elsewhere. This five-item scale has a reliability of .60. It correlates only .16 with Type A Sanctions, even though both scales would be elevated if a teacher frequently had to make some kind of response to misconduct. Useful information about the construct validity of the two sanctions scales can be obtained by examining their correlations with responses to a question about home-based reinforcers (Table 11). The use of home-based reinforcers to extend the range of rewards and punishments in the school appears to be a highly effective strategy (Barth, 1979; Atkeson & Forehand, 1979). Scores on the Type B scale correlate .35 with responses to this item, whereas those on Type A correlate only .07 (n.s.).

Victimization. As one way to measure the amount of "delinquent" behavior in a school, teachers are asked about their experiences of personal victimization. In the aggregate, these reports may be taken as an indicator of the amount of disruption in the school. An eight-item scale, asking about events ranging from obscene remarks or gestures to physical attack, has a reliability of .67.

Classroom Disruption. A second way to assess the level of student misbehavior experienced by a teacher is provided by a two-item classroom disruption scale. It asks to what degree

classroom disruption interferes with teaching, and how much of the teacher's time is devoted to coping with disruptive students. Its reliability is .70.

Low Expectations. A labeling theory perspective implies that teacher expectations for student performance may become incorporated into the student's self-concept and result in misconduct or poor academic performance. To provide a measure of this variable, a two-item Low Expectations Scale asks teachers to judge what percentage of their students are of low ability and have "behavior problems." The scale has a reliability of .57. It correlates -.24 with Job Satisfaction and .43 with Classroom Disruption.

Openness to Student Suggestions. This two-item scale has very low reliability. It asks, for example, how often teachers change their lesson plans to accomodate student suggestions. It was intended to provide an index of teacher responsiveness in order to assess the effectiveness of training programs or other interventions with teachers.

Professional Development. Eight items form a scale measuring the extensiveness of recent continuing education or in-service learning. This scale, with reliability .74, is for use in documenting the implementation of training components. It also helps to lend evidence of construct validity to other teacher measures. The correlations in Table 10 generally accord with the interpretation that teachers scoring high in professional development are more satisfied, interact more with students, and are more open to student suggestions.

Non-Authoritarian Attitudes. Intended in part to measure sympathetic attitudes (as one way to get at the "caring, competent teacher" constellation), a measure of punitive moralism is included. To earn a high score on this scale, a teacher rejects such items as, "A few pupils are just young hoodlums and should be treated accordingly." This three-item scale has a reliability of .54.

The item content of all teacher-level scales is shown in Appendix Table B.

Measures of School Climate

As discussed earlier, the assessment of school climates is fundamentally different from the measurement of individuals. Whereas individual differences are the entire point of measurement at the individual level, these differences are "error" or "noise" in the assessment of an environment based on the reports of its individual inhabitants. Consequently, environments are sometimes characterized by aggregated or averaged reports of individuals. We have constructed climate scales based on such aggregated reports, and sometimes describe climates using averaged characteristics of

individuals (cf. Astin & Holland, 1961). For climate scales, reports are first averaged, and then item analyses proceed based on school means for the items.

Measures of Climate Based on Student Reports

Community Crime. This is a three-item scale based on averaged responses to questions about whether there are gangs in the student's neighborhood, whether the gangs try to get the student to join and whether the student's parents were robbed in the last year. This scale may be useful in describing the community context of the school (cf. National Institute of Education, 1978). It has a reliability of .57. (An "outlier," i.e., a school with extreme values on many items, was dropped from the sample in these and subsequent item analyses reported here, possibly deflating most reliabilities somewhat.)

Gangs in School. This scale is composed of averaged responses to questions about whether there are gangs in the school and, if so, how much trouble they cause. The reliability of this scale is .80.

Safety. This is a 13-item scale asking if students stay away from any of a list of places in the school. It also asks if students feel safe at school, or if they fear someone will hurt them at school or on the way to school. It resembles what was called "School Climate" in the Schools Initiative Evaluation (Grant et al., 1979). Its reliability coefficient is .92.

Disruption. This four-item scale is based on averaged responses to questions about the students' having to fight to protect themselves, seeing teachers threatened or attacked, and being in classes that were totally stopped by disruptive students. It has a reliability of .42.

Individualized Instruction. This scale is an attempt to use student reports as evidence about the level of individualized instruction characterizing the school as a whole. Individualized instruction, as usually construed, involves the development of individual learning plans, rewards for improvement over past levels of performance, and a pace of instruction suited to the individual. Two aspects of this conception are incorporated in this measure--students' reports that they have individual learning plans, and reports that they can work at their own speed in class. The reliability coefficient is .58.

Student Disrespect. One theoretical perspective (Greenberg, 1977) assumes that delinquency is in part a result of a special status accorded youth, one which isolates them from meaningful adult roles and subjects them to degrading interpersonal exchanges to which

adults would not be subjected. This scale is intended to assess the degree to which students feel that a school environment as a whole either degrades them or treats them with dignity. A low score could indicate that students feel they are treated with dignity. The items are, "Students are treated like children here;" "Teachers treat students with respect;" and "Teachers do things to make students feel put down." Its reliability coefficient is .78.

Student-Teacher Interaction. This scale aims to assess the degree of out-of-class positive social interaction with teachers, from the students' point of view. It is based on the averaged responses to two items: "I talk to some of my teachers about things other than schoolwork;" and, "Teachers help me with schoolwork outside of class." Its reliability is .60.

Planning and Action. This scale is intended to assess, from the point of view of the students, the degree to which schools engage in experimenting and problem-solving, or the degree to which they resist change. It is composed of the following three aggregated items: "It is hard to change the way things are done in this school"; "The teachers and principal in this school make plans to solve problems"; and "This school hardly ever tries anything new." It has a reliability coefficient of .65.

Fairness. Evidence is accumulating that the degree to which students perceive a school's rules as fair and clear is associated with the degree of orderliness of the school (National Institute of Education, 1978; Gottfredson & Daiger, 1979). Consequently, scales designed to assess these constructs were developed. Fairness is a three-item aggregate-level scale based on student reports that the rules are fair, that the punishment for breaking rules is the same for everyone, and that the principal is fair. It has a reliability of .62.

Clarity. Intended to measure the clarity of school rules from the point of view of the school's students, this scale is composed of questions asking whether everyone knows what the rules are, whether teachers let the students know what is expected, whether the principal is firm. This four-item scale has a reliability coefficient of .64.

Student Influence. It is often assumed that student influence on the way a school is run may lead to a number of positive outcomes. This six-item scale is intended to assess how much influence students have in their schools. Sample items include: "Students have little say in how the school is run"; "Students have helped to make the school rules"; and "Students are seldom asked to help solve a problem the school is having." The scale's reliability is .62.

Grouping. This scale assesses the students' perceptions of grouping, or segregation of students with special characteristics within the school. It is composed of the following three items: "Students of different races usually end up in different classes"; and, "This school has special classes for slow learners"; and, "There are special classes for trouble makers." Its reliability is .55.

The correlations among the scales, shown in Table 13, are not low enough to imply that each scale measures an important independent dimension of school climate. In particular, scales 6 through 11 show considerable redundancy, implying that they should probably be interpreted as a group. Small differences in elevation among these scales should be interpreted with caution. The item content of these scales is summarized in Appendix Table C.

It seems appropriate to reduce this set of a priori scales through combination. This could result in producing the information with fewer, longer, and more reliable measures.

Climate Scales Based on Teacher Reports

An alternative perspective on the climate of a school is provided by the reports of teachers. Accordingly, seven climate scales were constructed from the teacher questionnaire, using averaged teacher responses about their school. The item content of these seven scales is presented in Appendix Table D. Their reliabilities are shown in Table 15.

Resources for Instruction. This scale is intended to measure relative levels of resources (equipment, materials, learning opportunities) available in the school. It contains items asking about teaching supplies, space, extra-school settings used for instruction, and timeliness of availability of resources. This four-item scale has a reliability of .86.

Involvement of Parents and Community. A goal of the Alternative Education Program is to increase the use of community and family resources by schools as a structural school improvement. This scale seeks to assess parent and community involvement according to aggregate teacher reports. It asks about parent influence on policies or practices, direct parent assistance, relations between parents and teachers, and community receptiveness. The six-item scale has a reliability of .80.

Teaching Staff Commitment. Some evidence suggests that the commitment of an organization's staff is related to project implementation (Grant et al., 1979; Berman & McLaughlin, 1976).

Accordingly, a two-item scale to assess staff commitment was included. Its reliability is .82.

Staff Morale. As with commitment, morale is sometimes suggested as a concomitant of success in implementing innovations, and it is an important characteristic of an organization in its own right. An 11-item scale containing items such as, "Our problems in this school are so big that it is unrealistic to expect teachers to make much of a dent in them;" and "(Is the teaching faculty) frustrated?" Its reliability is .90.

Planning and Action. Presumably, organizations engaging in systematic planning and that are open to change are most likely to successfully implement innovations. Based on this assumption, we constructed a nine-item scale to assess planning and action. It asks, "How often do you work on a planning committee with other teachers?" "(Is the principal) progressive?" "(Is the teaching faculty) open to change?" Its reliability is .87.

Smooth Administration. Our earlier research (Gottfredson & Daiger, 1979) suggests that the way a school is run is important in understanding its climate and in preventing school disruption. To the best of our knowledge, detailed studies of school administration tend to focus on the personal characteristics of administrators (e.g., Miner, 1967), or are ethnographic or observational accounts of the typical activities of administrators. Here we wished to assess the perceptions of administrative style and procedures from the point of view of the body of teachers who experience them. Accordingly, we constructed a 12-item scale. Typical items are: "Simple, non-time consuming procedures exist for the acquisition and use of resources;" "There is little teacher-administration tension in this school." "(The principal is) open." In a sense this scale represents a global rating of the positiveness with which teachers view the schools's administration, although the item content focuses on both principal behavior and some probable practical consequences of that behavior. Its reliability is .92.

Individualized Instruction and Grading. The Alternative Education Program seeks to create structural changes in schools to increase individualized instruction, and this intervention is planned by several of the action projects. Accordingly, this four-item scale aims to measure individualized instruction by asking if individualized learning plans are used, and if grading is based on improvement versus "the curve."

School Race Relations. This brief two-item measure asks about race relations from the teacher point of view. It asks how well different groups get along. Its reliability is .77.

Interaction with Students. The Alternative Education Program assumes that "caring competent teachers" will foster prosocial outcomes and prevent delinquency, and several action projects aim to alter teacher-student relations. Interpreting what "caring and competent" means is difficult, but as one way to get at this constellation we created an Interaction with Students Scale. This six-item index asks about the frequency of teacher interaction with students and about how well students and teachers get along. Its reliability is .80.

Integration vs. Segregation by Ability or Conduct. This scale is also included to measure an aspect of project implementation sought by the Alternative Education Program: the avoidance of tracking or isolation. The six-item scale contains items such as: "Students of mixed ability work together in small groups in my class;" "This school has special classes for slow learners;" and "In this school there are special classes for students who repeatedly misbehave." Its reliability is .55, and the appropriate interpretation of the scale is unclear. Opinions differ about the wisdom of homogeneous vs. heterogeneous grouping according to student conduct or academic performance, although the current climate, and some evidence (Slavin, 1980), implies that heterogeneous grouping can have some virtue.

Student Influence. Student participation in school decision making is one of the major structural elements the Alternative Education Program wants to create through the action projects. The assumption apparently is that student influence will help to create other beneficial structural changes, or it may contribute to decreased alienation or sense of powerlessness. Measures of student influence used in previous studies (National Institute of Education, 1978; Gottfredson & Daiger, 1979) assessed a limited range of influence, and certainly do not assess the kinds of student influence possible. Therefore, although based on the scale used earlier by Gottfredson & Daiger (1979), this scale is expanded somewhat (to five items). Sample questions are "I often change my lesson plans based on student suggestions;" and "Teachers and their students work together to make rules governing behavior in the classroom." The scale has a reliability coefficient of .81.

Professional Development. This scale is the aggregate-level counterpart of the individual-level teacher Professional Development Scale. Because of the way it is constructed, it is of use only for characterizing the level of professional development activities for the school as a whole. The

eight-item scale has a reliability of .86.

Perceptions of Disruption or Lack of Safety. Intended to be one measure of delinquent or disruptive behavior in school, this 12-item scale asks about time spent coping with disruptive behavior and about perceptions of safety. It is highly reliable (.93)

Success Opportunities. Many critics of traditionally structured schools (e.g., Howard, 1978) argue that schools are "rigged" against low achieving students: Many students experience only failure in school. This scale is an attempt to collect teacher opinions about the extent to which their schools are "unrigged" by providing success opportunities. It asks whether any students can earn high marks, and whether students can get special privileges for their performance. We have some doubts about the meaning of this scale: It is heterogeneous in content as well as brief, and it has modest reliability (.60). But it has some interest and may be of some utility.

Use of Grades as a Sanction. The use of grades as a response to misconduct is correlated with school disruption rates (Gottfredson & Daiger, 1979). On the face of it, this also appears to be a poor practice because it makes the grading and sanctioning process ambiguous. A two-item index uses teacher reports to characterize the extent of this practice in schools. It has a reliability of .84.

Interpreting Scores for Schools

One way of interpreting scores on all these scales is by using a profile sheet that enables the examination of the standing of a school relative to a "norm group." Profile sheets have been prepared that do just this, using the schools in the School Action Effectiveness Study as the norm group.<2> Remember that this is not a representative sample of schools. In general, these schools are included in the sample at least in part because they indicated in their applications that their problems of crime, dropout, and nonattendance are relatively severe. Nevertheless, the group does provide some basis for comparison.

Profile sheets are used to plot t-scores. This means that the "average" school would have a score of 50, and that the standard deviation is 10. As an aid to interpretation, a bar on the profiles shows the range of the middle 50% of the scores for each scale. This is especially useful when distributions of scores are skewed, or when they are flatter or more peaked than in a normal distribution, because the bar acts as an easy guide in interpreting the school's score.

For many of these measures, scores are not very reliable. Remember to take

reliability into account when making interpretations. In addition, "reliability" applies to scores in general, not the score for any particular school. When only a small number of persons reported about a school's climate, a score may contain lots of error. Similarly, if response rates were low, the score may contain bias. This is because students who completed the questionnaire may not accurately represent the schools' students in general. Do not make much of small differences in scores in these profiles, and be skeptical about profiles based on small samples. An interim rule of thumb to follow for profiles based on 50 or more questionnaires is to ignore the differences of five t-score units or less. When fewer than 50 questionnaires contribute to a school's score, even larger differences should be ignored.

If several sources of information converge in suggesting the same interpretation, the results are worthy of careful consideration. But remember, a survey like this is only one source of information about these schools. No survey or set of scales can provide a magic picture of an organization's environment.

The Uses of Scores for Individuals

All individual-level scores are confidential and are used for research purposes only. (In future applications, similar scales could be used in counseling or for diagnostic purposes. They can not be used in those ways in this Program because of the assurances of confidentiality given to those who completed the instruments on which scores are based.) Elsewhere in this report, and in subsequent reports to be issued by the SAES, these scores are used in detailed examinations of the effects of project components. When individual-level scores are aggregated to the school level, they can be used to describe schools in terms of the characteristics of their studentries. In profiles we make available to project directors, these aggregated scores are also presented as t-scores, where 50 is the mean aggregated score for schools, and 10 is the school standard deviation.

The Utility of Information for Project Managers

In workshops conducted in August, 1981, school profiles were made available to project directors. These profiles provide assessments of schools useful for diagnostic and prescriptive purposes. The efforts of thousands of students and teachers in completing these surveys will go partly to waste if this information is not used in project planning. We earnestly hoped that this information would be used, and are gratified that several projects have made extensive use of this information in renewed project planning.

Similarly, interim feedback we have provided to project directors on the characteristics of their clientele (in summary form), and about the effectiveness of their interventions based on the statistical analyses of individual scales is intended to be used in refining interventions. No one expects to see dramatic effects of projects in their developmental stages. Projects will increase in effectiveness largely by using the information provided by this interim feedback.

References

Astin, A. W., & Holland, J. L. The environmental assessment technique: A way to measure college environments. Journal of Educational Psychology, 1961, 52, 308-316.

Atkeson, B. M., & Forehand, R. Home-based reinforcement programs designed to modify classroom behavior: A review and methodological evaluation. Psychological Bulletin, 1979, 86, 1298-1308.

Bachman, J. G., O'Malley, P. M., & Johnston, J. Adolescence to adulthood: Change and stability in the lives of young men. Ann Arbor, Mich.: Institute for Survey Research, 1978.

Bandura, A. Social learning theory. Morristown, N.J.: General Learning Press, 1971.

Barth, R. Home-based reinforcement of school behavior: A review and analysis. Review of Educational Research, 1979, 49, 436-458.

Berman, P., & McLaughlin, M. W. Implementation of educational innovation. Educational Forum, 1976, 40, 345-370.

Bielby, W. T., Hauser, R. M., & Featherman, D. L. Response errors of nonblack males in models of the stratification process. American Journal of Sociology, 1977, 82, 1242-1288.

Coopersmith, S. The antecedents of self-esteem. San Francisco: Freeman, 1967.

Duncan, O. D. A socioeconomic index for all occupations. In A. J. Reiss, Jr. (Ed.), Occupations and social status. New York: Free Press, 1961.

Elliott, D. S., & Ageton, S. Reconciling differences in estimates of delinquency. American Sociological Reveiw, 1980, 45, 95-110.

Empey, L. T. American delinquency. Homewood, Ill.: Dorsey, 1978.

Empey, L. T., & Erickson, M. L. Hidden delinquency and social status. Social Forces, 1966, 44, 546-554.

Fox, R., and associates. School climate improvement. Tulsa: CADRE, 1974.

Gottfredson, D. C. Black-white differences in educational attainment. American Sociological Review, 1981, 46, 542-557. (a)

Gottfredson, D. C. Modeling delinquency: More inconclusive evidence. Paper presented at the annual meeting of the American Educational Research Association, 1981. (b)

Gottfredson, G. D. Schooling and delinquency. In S. E. Martin, L. B. Sechrest, & R. Redner (Eds.), New directions in the rehabilitation of criminal offenders. Washington, D.C.: National Academy Press, 1981. (a)

Gottfredson, G. D. Schooling and delinquency prevention: Some practical ideas for educators, parents, program developers, and researchers (Report No. 304). Baltimore: Johns Hopkins University, Center for Social Organization of Schools, 1981. (b)

Gottfredson, G. D., & Daiger, D. C. Disruption in six hundred schools: The social ecology of school victimization. (Report No. 281). Baltimore: Johns Hopkins University, Center for Social Organization of Schools, 1979.

Gottfredson, G. D., Joffe, R. D., & Gottfredson, D. C. Measuring victimization and the explanation of school disruption (Report No. 306). Baltimore: Johns Hopkins University, Center for Social Organization of Schools, 1981.

Gough, H. G. Manual for the California Psychological Inventory. Palo Alto, Calif.: Consulting Psychologists Press, 1964.

Grant, J., Grant, J. D., Daniels, D., Neto, V., & Yamasaki, C. The school team approach phase I evaluation (Prepared under grants no. 77-NI-99-0012 and 78-JN-AX-0016 from the National Institute for Juvenile Justice and Delinquency Prevention). San Rafael, Calif.: Social Action Research Center, 1979.

Greenberg, D. F. Delinquency and the age structure of society. Contemporary Crisis, 1977, 1, 189-223.

Greenberger, E., Campbell, P., Sorensen, A. B., & O'Connor, J. Toward the measurement of psychosocial maturity (Report No. 110). Baltimore: Johns Hopkins University, Center for Social Organization of Schools, 1971. (ERIC No. ED 052 262)

Hindelang, M. J., Hirschi, T., & Weis, J. G. Measuring delinquency. Beverly Hills, Calif.: Sage, 1981.

Hirschi, T. Causes of delinquency. Berkeley: University of California Press, 1969.

Holland, J. L., & Baird, L. L. An interpersonal competency scale. Educational and Psychological Measurement, 1968, 28, 503-510.

Hoppock, R. Job satisfaction. New York: Harper, 1935.

Howard, E. R. School discipline desk book. West Nyack, N.Y.: Parker, 1978.

Jencks, C. Who gets ahead? The determinants of economic success in America. New York: Basic Books, 1977.

Lemert, E. M. Human deviance, social problems, and social control. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

McPartland, J. M., & McDill, E. L. Research on crime in schools. In J. M. McPartland & E. L. McDill (Eds.), Violence in schools. Lexington, Mass.: Lexington, 1977.

Miller, W. B. Lower class culture as a generating milieu of gang delinquency. Journal of Social Issues, 1958, 14, 5-19.

Miner, J. B. The school administrator and organizational character. Eugene, Oregon: Center for Advanced Study of Educational Administration, 1967.

Moos, R. H. A typology of junior and high school classrooms. American Educational Research Journal, 1978, 15, 53-66.

Moos, R. H., & Trickett, E. J. Manual for the Classroom Climate Scale. Palo Alto, Calif.: Consulting Psychologists Press, 1974.

National Institute of Education. Violent schools--Safe schools: The safe school study report to Congress. Washington, D.C.: Author, 1978.

Nye, I. F. Family relationships and delinquent behavior. New York: Wiley, 1958.

Otto, L. B. Social integration and the status-attainment process. American Journal of Sociology, 1976, 81, 1360-1383.

Porter, L. W., & Lawler, E. E. III. Managerial attitudes and performance. Homewood, Ill.: Dorsey, 1968.

Richards, J. M., Jr. The social climate scales. In O. K. Buros (Ed.), The eighth mental measurement yearbook. Highland Park, Ill.: Gryphon, 1978.

Robinson, J. P., Athanasiou, R., & Head, K. B. Measures of occupational attitudes and occupational characteristics. Ann Arbor: Institute for Social Research, 1969.

Robinson, J. P., & Shaver, P. R. Measures of social psychological attitudes (rev. ed.). Ann Arbor: Institute for Social Research, 1973.

Rosenberg, M. Society and the adolescent self-image. Princeton: Princeton University Press, 1965.

Sewell, W., Haller, A., & Portes, A. The educational and early occupational attainment process. American Sociological Review, 1969, 34, 82-92.

Slavin, R. E. Cooperative learning in teams: State of the art. Educational Psychologist, 1980, 15, 93-111.

Srole, L. Social integration and certain corollaries. American Sociological Review, 1956, 21, 709-716.

Stanley, J. C. Reliability. In R. L. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council on Education, 1971.

Stern, G. People in context: Measuring person-environment congruence in education and industry. New York: Wiley, 1970.

Sutherland, E. H., & Cressey, D. R. Principles of criminology (5th ed.). Philadelphia: Lippincott, 1955.

Thorndike, R. L. (Ed.). Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Tittle, C. R., & Villemez, W. J. Social class and criminality. Social Forces, 1977, 56, 474-502.

Treiman, D. J. Occupational prestige in comparative perspective. New York: Academic Press, 1977.

Footnotes

1. For more thorough discussion see Throndike (1971).

2. One school with extremely deviant scores on a number of scales was excluded from the norm group.

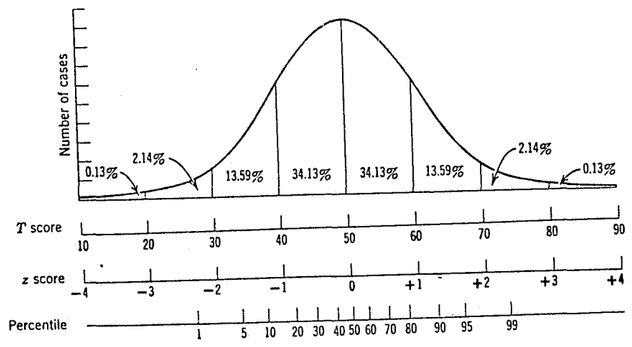


Figure 1. The interpretation of standard scores in a normal distribution

Table 1
Selected Characteristics of National Samples of Students and the School Action Effectiveness Study Sample

Ethnic Identification	Elementary & Secondary Enrollment[a]	High School Seniors[b]	SAES Sample[c]
American Indian	0.8	1.1	2.0
Asian-American	1.2	0.7	1.1
Spanish-American	6.4	3.2	25.4
Black	15.5	11.7	43.7
White	76.0	80.4	25.6
Other	--	2.9	2.1

Father Present?	Americans Aged 18 and Under	High School Seniors[e]	SAES[c]
Yes (%)	79.0	81.7	56.8

Type of Personal Victimization	Percentage Reporting in Month: National Sample[f]	SAES[g]
Physical attack	9.8	10.3
Robberies of more than $1	3.4	6.5
Personal theft of more than $1	27.6	23.0

Self-reported Delinquent Behavior	Percentage Reporting at Least One, Past Year: National Sample[g]	SAES[c]
Destroy or damage school property	11.4	11.1
Stole something worth more than $50	2.6	6.7
Carried a hidden weapon	6.3	12.0
Gang fights	7.9	10.4
Hit teacher	6.6	9.7
Hit students	5.0	45.4
Joy riding	4.5	6.5
Break in	2.5	5.1

a. Fall 1976 OCR data
b. Monitoring the Future, 1976
c. Unweighted
d. CPR
e. Monitoring the Future, 1976
f. Safe School Study
g. National Youth Survey, 1978

Table 2

Reliability Coefficients (Alpha) for Individual-Level Student Scales by Gender and Ethnic Self-Identification

Scale	Spanish-Americans Construction		Spanish-Americans Holdout		Blacks Construction		Blacks Holdout		Whites Construction		Whites Holdout		American Indian		Asian American		Total Sample	Number of items
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female		
Family background																		
Parental education[a]	--	--	73	72	--	--	77	78	--	--	84	72	86	68	66	62	76	2
Parental emphasis on education	47	44	47	39	42	38	42	41	48	56	43	44	45	52	50	55	46	4
Maternal role model (negative)	39	39	20	42	35	33	46	35	10	31	09	25	09	47	61	53	36	5
Maternal role model (positive)	46	49	46	44	37	45	53	43	55	51	49	48	32	62	30	58	46	4
Paternal role model (negative)	47	57	56	57	53	51	53	58	56	48	54	62	48	71	75	22	55	5
Paternal role model (positive)	67	68	74	72	72	69	77	71	70	75	50	75	55	75	57	73	71	6
Social Relations																		
Attachment to parents	52	66	50	61	52	59	50	60	72	81	66	78	68	77	73	73	61	6
Negative peer influence	65	60	61	56	55	57	60	53	71	73	72	69	60	62	74	60	64	9
Parental supervision	27	23	27	14	11	17	26	17	31	29	31	40	28	b	15	54	27	2
Attitudes and Social Development																		
Alienation	40	39	37	45	35	44	33	39	56	53	46	56	39	44	06	55	44	4
Attachment to school	75	70	74	70	72	71	67	67	81	81	79	79	79	75	81	72	75	10
Belief in rules	46	42	53	42	51	43	44	42	59	52	54	49	70	56	51	67	50	6
Interpersonal competency	47	43	40	45	46	36	44	30	44	37	46	34	40	44	41	32	42	5
Involvement	63	64	64	64	63	65	66	62	61	59	62	62	59	63	80	52	64	15
Positive self-concept	56	51	60	58	52	64	59	61	73	69	70	69	71	67	65	78	63	12
Practical knowledge	67	66	70	70	71	67	71	72	77	76	74	74	69	76	63	54	71	7
Rebellious autonomy	45	52	42	59	31	45	34	42	45	53	41	49	59	60	42	40	46	3
Behavior																		
School effort	39	44	37	37	53	48	51	50	67	65	62	63	52	47	62	65	51	5
School nonattendance	63	64	64	61	57	61	65	59	74	69	69	71	62	60	74	29	66	2
Self-reported delinquency (total)	81	79	87	77	81	71	85	76	88	84	87	83	85	81	93	63	84	19
Self-reported drug use	71	71	72	67	72	66	73	65	79	79	73	80	78	80	53	71	75	5
Self-reported serious delinquency	72	74	80	61	75	59	80	69	85	79	82	73	74	76	92	c	79	11
School Experiences																		
School punishments	51	58	59	58	49	50	48	44	57	54	51	46	47	32	64	70	54	4
School rewards	50	53	57	56	55	53	54	50	52	54	51	47	49	64	66	52	54	6
Victimization	71	67	72	64	68	65	72	56	63	61	70	61	66	48	82	69	68	7
Validity																		
Invalidity	45	36	40	40	51	41	52	46	48	33	50	36	53	02	55	57	47	5

Note: Decimals are omitted.

[a] Reliabilities estimated on entire sample, no separate calculations were made for construction and hold-out samples.

[b] Reliability too low to estimate.

[c] Too few cases with complete data to estimate reliability.

Table 3

Means and Standard Deviations of Raw Score Individual-Level Student Scales

Scale	Spanish-American						Black						White						American Indian						Asian American					
	Male			Female			Male			Female			Male			Female			Male			Female			Male			Female		
	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N
Family background																														
Parental education	3.20	2.57	536	2.56	2.30	616	4.84	2.15	458	4.76	2.23	1187	5.22	2.32	413	5.18	2.33	430	4.30	1.96	53	4.40	2.19	53	5.28	2.55	21	4.56	2.68	16
Parental emphasis on education	-.36	2.52	439	-.29	2.41	456	.25	2.25	303	.18	2.20	365	-.59	2.61	138	-.32	2.54	162	-1.62	2.80	42	.14	2.60	50	-.62	2.88	31	-.75	2.72	20
Maternal role model (negative)	.31	.57	472	.35	.68	509	.57	.87	342	.55	.79	408	.66	.75	146	.64	.79	162	.61	.72	51	.78	.98	54	.59	1.01	32	.29	.64	21
Maternal role model (positive)	4.94	1.13	442	4.78	1.13	469	4.88	1.20	332	4.93	1.10	388	4.66	1.26	148	4.76	1.22	160	3.85	1.05	48	4.96	1.31	49	4.87	1.09	31	3.80	1.28	20
Paternal role model (negative)	.77	1.05	441	.84	1.10	461	.78	1.04	308	1.03	1.19	361	.93	1.12	137	1.03	1.23	160	1.22	1.15	46	1.00	1.35	43	1.06	1.44	31	.68	.84	22
Paternal role model (positive)	4.88	1.53	437	4.54	1.58	442	4.85	1.60	314	4.56	1.57	362	5.09	1.12	138	4.60	1.56	159	4.85	1.23	40	4.18	1.77	45	4.67	1.40	30	4.45	1.60	22
Social relations																														
Attachment to parents	.85	3.04	399	.47	3.37	442	.18	3.16	567	.22	3.25	702	-.34	3.49	186	-.81	4.18	230	-1.32	3.91	48	-.77	4.11	50	-.60	4.26	36	-1.47	4.44	24
Negative peer influence	.44	4.65	524	-1.17	3.72	581	.39	4.62	850	-1.10	3.60	967	1.69	5.40	467	-.01	4.63	561	2.99	5.13	81	.19	4.46	84	.91	5.88	43	-.23	4.27	33
Parental supervision	1.51	.65	464	1.70	.53	523	1.25	.73	689	1.55	.62	853	1.30	.73	434	1.51	.67	496	1.09	.76	78	1.51	.60	73	1.20	.72	51	1.42	.76	31
Attitudes and social development																														
Alienation	1.37	1.11	445	1.25	1.12	520	1.35	1.08	719	1.31	1.10	856	1.42	1.18	485	1.31	1.22	572	1.63	1.17	73	1.27	1.13	70	1.30	.95	50	1.41	1.23	34
Attachment to school	-.46	5.60	427	1.17	4.90	514	.37	4.95	692	1.40	4.61	820	-1.71	6.31	463	-.28	5.84	534	-1.89	6.40	68	1.07	5.04	68	-.23	6.19	48	1.75	4.70	32
Belief in rules (-)	2.14	1.53	415	1.67	1.31	492	2.36	1.42	652	1.93	1.33	787	2.21	1.53	325	1.56	1.32	396	2.48	1.82	64	2.27	1.52	56	2.56	1.58	41	2.03	1.65	30
Interpersonal competency	3.70	1.18	427	3.78	1.18	506	3.86	1.14	678	4.05	.96	815	3.84	1.17	333	4.08	1.00	409	3.92	1.12	59	3.82	1.18	61	3.78	1.15	46	3.87	1.02	31
Involvement	-1.61	5.65	496	-1.38	5.74	533	.06	6.17	761	.77	6.06	849	-1.76	5.25	403	-.48	5.76	443	-1.19	5.61	67	.79	6.25	76	2.15	8.33	41	2.44	5.76	29
Positive self-concept	-1.24	5.37	281	-.06	4.98	342	.39	5.03	426	2.17	4.42	549	-.22	5.69	271	.84	4.99	330	.51	5.52	37	.35	5.32	45	-2.04	6.10	34	2.24	5.02	24
Practical knowledge	11.91	2.94	451	11.58	3.13	524	11.95	2.95	769	11.80	3.03	894	11.88	3.08	507	11.89	3.10	597	11.56	3.13	72	11.87	3.18	75	12.42	2.48	52	12.12	2.41	34
Rebellious autonomy	1.76	1.00	446	1.66	1.10	518	1.84	.95	674	1.81	.99	811	2.09	.93	351	2.18	.94	409	2.18	.98	65	2.07	1.03	61	1.75	1.01	44	1.79	.99	33
Behavior																														
School effort	7.04	1.98	557	7.63	1.91	585	7.74	1.87	889	8.37	1.78	996	7.31	2.13	426	8.16	1.93	492	7.31	1.98	80	8.13	1.83	78	7.52	2.20	42	8.58	1.87	36
School non-attendance	1.73	1.96	620	1.43	1.72	636	1.12	1.67	1081	.85	1.39	1121	1.50	2.03	623	1.34	1.95	700	1.64	1.98	104	1.71	1.96	99	1.54	2.18	63	1.07	1.49	41
Self-reported delinquency (total)	2.28	3.20	326	1.18	1.92	394	2.68	3.21	271	2.17	2.39	295	3.57	3.73	178	2.23	2.80	218	5.02	4.00	44	4.72	3.47	43	3.93	5.14	14	.90	1.10	10
Self-reported drug use	.84	1.21	356	.65	1.04	432	.93	1.29	489	.95	1.18	557	1.45	1.46	420	1.74	1.65	491	1.98	1.66	65	2.28	1.75	60	.91	1.08	23	.50	.80	12
Self-reported serious delinquency	.98	1.79	345	.33	.85	424	1.39	2.09	456	.49	1.12	525	1.45	2.23	206	.58	1.27	248	1.94	2.20	49	1.19	1.81	54	2.60	3.52	15	--	--	--
School experiences																														
School punishments	.19	2.77	545	-.86	2.14	610	.77	2.78	865	-.16	2.35	985	.08	2.58	580	-.64	2.14	669	.31	2.60	91	-.55	2.03	82	.39	2.96	55	-.31	2.75	40
School rewards	.42	3.47	498	.58	3.38	570	.48	3.34	764	.13	3.20	906	-1.32	3.00	402	-.89	2.95	469	-.04	3.24	76	.61	3.69	70	.21	3.62	46	.76	3.10	37
Victimization	1.04	1.49	568	.64	1.10	618	1.22	1.59	891	.82	1.13	992	1.00	1.42	541	.69	1.09	618	1.25	1.52	89	.80	1.04	83	1.32	1.87	57	1.22	1.53	41
Validity																														
Invalidity	1.24	1.14	470	.94	1.04	532	1.09	1.19	754	.77	.99	874	.65	.96	513	.45	.75	610	.87	1.09	77	.59	.71	69	1.29	1.29	51	.71	1.06	34

Table 4

Correlations of Student Scales and Selected Other Characteristics with Self-Report Indexes of Student Behavior

Scale or Characteristic	School Effort	Self-Reported Delinquency		
		Total	Drug	Serious
Parental education	15*	06	03	06
Parental emphasis on education	19*	-22*	-22*	-17*
Maternal role model (negative)	-11*	22*	17*	19*
Maternal role model (positive)	11*	-07*	-05	-08*
Paternal role model (negative)	-10*	16*	15*	13*
Paternal role model (positive)	12*	-07*	-09*	-05
Attachment to parents	22*	-26*	-31*	-21*
Negative peer influence	-31*	50*	41*	50*
Parental supervision	14*	-27*	-22*	-24*
Alienation	-22*	20*	16*	20*
Attachment to school	35*	-34*	-30*	-34*
Nonbelief in rules	-22*	27*	20*	30*
Interpersonal competency	22*	-03	01	-08*
Involvement	15*	02	-06*	03
Positive self-concept	48*	-24*	-19*	-29*
Practical knowledge	07*	02	04	-01
Rebellious autonomy	-07*	18*	19*	13*
School punishments	-22*	30*	17*	30*
School rewards	15*	-09*	-14*	-04
Victimization	-15*	24*	09*	25*
Self-reported reading ability	20*	-02	-02	-05*
Self-reported school grades	39*	-11*	-12*	-14*
Age	-06*	08*	22*	03
Sex (male)	-17*	20*	01	25*

Note: Decimals omitted. Sample sizes are very large, so even small correlations are significant. Size rather than significance of these coefficients should be attended to.

*$p < .001$

Table 5

Correlations Among Student Measures of Family Background

Scale	Scale					
	1	2	3	4	5	6
1. Parental education	(76)	24	07	11	-14	17
2. Parental emphasis on educ.		(46)	-17	30	-20	31
3. Maternal role model (neg.)			(36)	-27	27	-09
4. Maternal role model (pos.)				(46)	-07	19
5. Paternal role model (neg.)					(55)	-40
6. Paternal role model (pos.)						(71)

Note: Reliabilities shown in diagonal. Decimals omitted.

Table 6

Correlations Among Student Measures of Social Relations

Scale	Scale		
	1	2	3
1. Attachment to parents	(61)	-26	21
2. Negative peer influence		(64)	-25
3. Parental supervision			(27)

Note: Reliabilities shown in diagonal. Decimals omitted.

Table 7

Correlations Among Measures of Student Attitudes and Social Development

Scale	Scale 1	2	3	4	5	6	7	8
1. Alienation	(44)	-53	27	-20	-07	-39	-07	19
2. Attachment to school		(75)	-34	27	13	41	09	-21
3. Non-belief in rules			(50)	-13	-02	-31	-01	24
4. Interpersonal competency				(42)	10	39	17	09
5. Involvement					(64)	11	12	-06
6. Positive self-concept						(63)	12	-07
7. Practical knowledge							(71)	03
8. Rebellious autonomy								(46)

Note: Reliability coefficients shown in diagonal cells. Decimals omitted.

Table 8

Correlation of Student Scales with School Experiences

Scale	Scale 1	2	3
1. School punishments	(54)	07	35
2. School rewards		(54)	16
3. Victimization			(68)

Note: Reliability coefficients shown in diagonal cells. Decimals are omitted.

Table 9

Reliability Coefficients for the Individual-Level Teacher Scales and Raw Score Item Statistics

Scale	Number of items	Construction Sample[a] Mean	SD	Alpha	Hold-out Sample[b] Mean	SD	Alpha
Prointegration attitude	4	11.53	2.81	.67	11.56	2.88	.69
Job satisfaction	3	8.45	1.64	.78	8.42	1.70	.80
Interaction with students	6	14.20	4.43	.69	13.79	4.20	.67
Type A sanctions	5	-.05	2.90	.52	.09	2.82	.47
Type B sanctions	5	13.81	3.08	.58	13.42	3.08	.60
Victimization	8	1.24	1.45	.67	1.23	1.45	.67
Classroom disruption	2	4.52	1.22	.70	4.60	1.38	.78
Low expectations	2	62.43	42.89	.53	65.06	44.47	.57
Openness to student suggestions	2	.05	1.64	.45	-.10	1.54	.38
Professional development	8	-.21	4.82	.76	-.39	4.67	.74
Nonauthoritarian attitude	3	7.57	2.21	.56	7.43	2.17	.54

[a] N's range from 555 to 642 due to nonresponse to some items.

[b] N's range from 555 to 643 due to nonresponse to some items.

Table 10

Correlations Among Individual-Level Teacher Scales
(N=1112-1265)

Scale	Prointeg.	Job Sat.	Interact.	Type A	Type B	Victim.	Disrup.	Low Exp.	Openness	Prof.Dvt.	Nonauth.
Prointegration attitude	--	06	12*	-09	04	-09	-10*	-07	15*	08	29*
Job satisfaction		--	28*	-13*	00	-20*	-34*	-24*	11*	28*	13*
Interaction with students			--	-02	11*	01	-15*	-12*	24*	28*	11*
Type A sanctions				--	16*	19*	29*	12*	-03	-03	-26*
Type B sanctions					--	08	12*	12*	09	17*	02
Victimization						--	38*	25*	-02	-07	-18*
Classroom disruption							--	43*	-02	-12*	-16*
Low expectations								--	01	05	-07
Openness to stud. sug.									--	19*	14*
Professional develop.										--	10*
Nonauthoritarian											--

Note: N's vary from 1112 to 1265 due to teacher nonresponse to some items.

* $p < .001$

Table 11

Correlations of Individual-Level Teacher Scales with Selected Teacher Characteristics

Scale	Sex (female)	Non-white	Use of home-based reinforcement	Item non-response index	Range of N
Prointegration attitudes	10*	21*	04	04	1204-1241
Job satisfaction	01	07	06	-08	1274-1315
Interaction with students	-05	13*	17*	06	1246-1289
Type A sanctions	-10*	-02	07	-01	1169-1206
Type B sanctions	15*	02	35*	22*[a]	1192-1230
Victimization	-15*	-18*	08	-11*	1175-1212
Classroom disruption	-04	-15*	12*	-02	1183-1223
Low expectations	04	-02	11*	02	1168-1205
Openness to student influence	06	09	09	02	1128-1164
Professional development	13*	27*	16*	10*	1190-1223
Nonauthoritarian attitude	15*	01	-04	08	1207-1245

Note: Decimals omitted

*$p < .001$

[a]High scorers on the scale measuring use of type B sanctions failed to respond to many items intended to measure type A sanctions ($r=.44$), suggesting that they were using the questionnaire as a checklist rather than carefully marking a response for each item.

Table 12

Reliabilities of the School Climate Measures Based on Aggregated Student Reports

	Alpha	Number of items
Community Crime	.57	3
Gangs in School	.80	2
Safety	.92	13
Disruption	.42	4
Individualized Instruction	.58	2
Student Disrespect	.78	3
Student-Teacher Interaction	.60	2
Planning and Action	.65	3
Fairness	.62	3
Clarity	.64	4
Student Influence	.62	6
Grouping	.55	3

Table 13

Correlations Among School Climate Measures Based on Aggregated Student Reports

Scale	Scale 2	3	4	5	6	7	8	9	10	11	12
1. Community Crime	.39*	-.24	.14	.24	-.03	-.22	-.08	-.18	-.16	-.04	.13
2. Gangs in School		-.01	-.02	-.12	.01	-.33	-.30	-.18	-.26	-.30	.18
3. Safety			-.43*	-.40*	-.13	-.02	-.03	-.06	.10	-.27	-.22
4. Disruption				.36*	.66*	-.36*	-.33	-.49*	-.16	-.20	.55*
5. Individualized Instruction					-.22	.01	-.02	.02	.02	.37*	.17
6. Student Disrespect						-.59*	-.39*	-.60*	-.30	-.53*	.33
7. Student-Teacher Interaction							.45*	.42*	.35*	.43*	-.28
8. Planning and Action								.50*	.68*	.62*	-.39*
9. Fairness									.45*	.56*	-.37*
10. Clarity										.48*	-.24
11. Student Influence											-.22
12. Grouping											

Note.--N's range from 52 to 65 schools, depending on the availability of items to score these scales.

Table 14

Reliabilities of the School Climate Measures
Based on Aggregate Teacher Reports

Scale	Alpha	Number of Items
Involvement of Parents and Community	.80	6
Individualized Instruction and Grading	.60	4
Resources for Instruction	.86	4
Integration vs. Segregation by Ability or Conduct	.55	6
School Race Relations	.77	2
Interaction with Students	.80	6
Teaching Staff Commitment	.82	2
Use of Grades as a Sanction	.84	2
Success Opportunities	.60	3
Staff Morale (vs. Alienation)	.90	11
Planning and Action	.87	10
Perceptions of Disruption	.93	13
Student Influence	.81	5
Smooth Administration	.92	12
Professional Development	.86	8

Note: N=48 to 50 schools (based on responses of over 1,100 teachers).

Table A

Item Content of Individual-Level Student Scales<a>

Parental Education

20. How far did your father (or stepfather) go in school?
21. How far did your mother (or stepmother) go in school?

Parental Emphasis on Educaton

23. Do your parents want you to go to college some day?
28. My parents keep close track of how well I am doing in school.
26. (Father) helps me with my homework.
27. (Mother) helps me with my homework.

Maternal Role Model (Negative)

26. (Mother) drinks too much.
26. (Mother) gets in trouble with the police.
26. (Mother) spends most of her money on herself.
26. (Mother) gets mad a lot.
26. (Mother) spends time with her friends away from the house.

Maternal Role Model (Positive)

26. (Mother) helps me with personal problems.
26. (Mother) is a hard worker.
26. (Mother) gives me money when I need it.
26. (Mother) goes to work every day.

Paternal Role Model (Negative)

26. (Father) drinks too much.
26. (Father) gets in trouble with the police.
26. (Father) spends most of his money on herself.
26. (Father) gets mad a lot.
26. (Father) spends time with his friends away from the house.

<a>Numbers indicate position of the item in the questionnaire.

Table A (cont.)

Paternal Role Model (Positive)

26. (Father) helps me with personal problems.
26. (Father) is a hard worker.
26. (Father) gives me money when I need it.
26. (Father) goes to work every day.

Attachment to Parents

29. How much do you want to be like the kind of person your mother (or stepmother) is?
30. How close do you feel to your parents (or guardians)?
31. How much do you want to be like the kind of person your father (or stepfather) is?
32. All in all, how much do you like your parents?
68. I would not care if my parents were a little disappointed in me.
69. I have lots of respect for my parents.

Negative Peer Influence

43. Most of my friends think getting good grades is important. (-)
43. Most of my friends think school is a pain.
43. My friends often try to get me to do things the teacher doesn't like.
44. (Best friend) is interested in school. (-)
44. (Best friend) attends classes regularly. (-)
44. (Best friend) plans to go to college. (-)
44. (Best friend) belongs to a gang.
44. (Best friend) gets in trouble with the police.
45. How many of your friends have been picked up by the police?

Parental Supervision

28. My parents almost always know where I am and what I am doing.
69. As far as my parents are concerned, I am pretty much free to come and go as I please.

Alienation

64. Teachers here care about the students. (-)
64. I feel like I belong in this school. (-)

Table A (cont.)

69. I feel no one really cares much about what happens to me.
70. These days I get the feeling that I'm just not a part of things.

Attachment to School

62. (How important is) what the teachers think about you?
63. (How do you feel about) this school?
63. (How do you feel about) the principal?
63. (How do you feel about) the classes you are taking?
63. (How do you feel about) the teacners?
63. (How do you feel about) the counselors?
64. I have lots of respect for my teachers.
64. This school makes me like to learn.
62. (How important is) tne grade you get at school?
70. In classes I am learning the things I need to know.

Belief in Rules

69. It is all right to get around the law if you can. (-)
69. People who leave things around deserve it if they get taken. (-)
70. Taking things from stores doesn't hurt anyone. (-)
70. It is O.K. to take advantage of a chump or a sucker. (-)
70. Teachers who get hassled by students usually had it coming. (-)
68. I do not have much to lose by causing trouble in school.

Interpersonal Competency

69. I have a clear picture of what I am like as a person.
69. I know how to get along with teachers.
69. If I want to, I can explain things well.
70. I find it easy to talk to all kinds of people.
70. My friends regard me as a person with good sense.

Involvement

34. Which of the following things have you spent time on this school term? (12-item list follows)
35. How much time, on the average, do you spend doing homework?
36. Did you do any work for pay last week, not counting work around the house? 37. Do you have a regular part-time or full-time job?

Table A (cont.)

Positive Self-Concept

7. How satisfied are you with the way you are doing in school?
69. Sometimes I think I am no good at all. (-)
70. I feel I do not have much to be proud of. (-)
70. I like myself.
9. (Other students see me as) a loser. (-)
68. I am the kind of person who will always be able to make it if I try.
9. (Other students see me as) a good student.
9. (Other students see me as) a trouble maker. (-)
9. (Other students see me as) successful.
68. My teachers think that I am a slow learner. (-)
68. I do not mind stealing from someone--that is just the kind of person I am. (-)
68. I am not the kind of person you would expect to get in trouble with the law.

Practical Knowledge

73. Do you know how to: (seven competency items)

Rebellious Autonomy

69. I don't like anybody telling me what to do.
70. Whether or not I spend time on homework is my own business.
70. I should not have to explain to anyone how I spend my money.

School Effort

8. Compared to other students, how hard do you work in school?
38. I turn my homework in on time.
38. My schoolwork is messy. (-)
38. I don't bother with homework or class assignments. (-)
38. If a teacher gives a lot of homework, I try to finish all of it.

Table A (cont.)

School Nonattendance

13. In the last four weeks, how many days did you cut class all day?
14. How often do you cut one or more of your classes?

Self-Reported Delinquency (Total)

46-49. Respondents mark "yes" or "no" to 19 kinds of behavior.

Self-Reported "Drug" Use

46-49. In the last year have you ...
...smoked cigarettes?
...drunk beer, wine, or "hard" liquor?
...smoked marijuana (grass, pot, ganja)?
...gone to school when you were drunk or high on some drugs?
...taken some other drugs?.

Self-Reported "Serious" Delinquency

46-49. In the last year have you ...
...purposely damaged or destroyed property belonging to a school?
...purposely damaged or destroyed other property not belonging to you, not counting family or school property?
...stolen or tried to steal something worth more than $50?
...carried a hidden weapon other than a plain pocket knife?
...been involved in gang fights?
...hit or threatened to hit a teacher or other adult at school?
...taken a car for a ride (or drive) without the owner's permission?
...used force or strong-arm methods to get money or things from a person?
...stolen or tried to steal things worth less than $50?
...stolen or tried to steal something at school, such as someone's coat from a classroom, locker, or cafeteria, or a book from the library?
...broken or tried to break into a building or car to steal something or just to look around?

Table A (cont.)

School Punishments

59. Did you have to stay after school as a punishment?
59. Did you get an extra assignment as a punishment?
59. Was your grade lowered on an assignment as a punishment?
59. Were you sent out of class for punishment?

Rewards (Note: Contains two items that should be removed.)

51. Students get to help other students.
58. Teachers say nice things about my classwork.
59. Did you get to do something special as a reward?
59. Did you win an award or prize because of your work in school?
59. Did you help win an award or a prize for your group or class because of your work in school?
61. Students who are well-behaved in this school get special treatment.

Victimization

60. Seven-item list. See questionnaire.

Invalidity

I have never disliked anyone. (T)
It is easy to get along with nasty people. (T)
I sometimes get angry. (F)
I like to have fun. (F)
I read several whole books every day. (T)

Table B

Individual-Level Teacher Scales

Pro-Integration Attitude

Most black students are better off in all-black schools. (-)
Most white students are better off in all-white schools. (-)
The amount of prejudice against minority groups in this country is greatly exaggerated. (-)
Students should not be bussed to achieve racial balance. (-)

Job Satisfaction

How do you like your job?
How much of the time do you feel satisfied with your job?
How much do you think you like your job compared with other people?

Interaction with Students

In the past two weeks have any students come to you to ask your advice on some prolbem they were haveing outside of class?
How often do you engage in the following activities with students:
- . . . tutoring individual students before or after school.
- . . . working with students on extracurricular activities.
- . . . taking students on field trips.
- . . . going to games, dances, and other student activities.
- . . . discussing students' personal problems with them.

Type A Sanctions

In your dealings with misbehaving students how often do you do the following things?
- . . . send them out of class.
- . . . use or threaten to use physical punishment.
- . . . lower their grades if misconduct is repeated.
- . . . reprimand the student in tne class.

When a student misbehaves in my class, I sometimes lower his or her grade.

Table B (cont.)

Type B Sanctions

In your dealings with misbehaving students how often do you do the following things?
- . . . give additional school work.
- . . . give privileges to increase positive behavior.
- . . . withdraw privileges for misconduct.
- . . . call a parent.
- . . . get help from a counselor.
- . . . refer the student to a special program.

Victimization

How many times in the past month have the following happened to you personally in this school?
- . . . damage to personal property worth less than $10.
- . . . damage to personal property worth more than $10.
- . . . theft of personal property worth less than $10.
- . . . theft of personal property worth more than $10.
- . . . was physically attacked but not seriously enough to see a doctor.
- . . . received obscene remarks or gestures from a student.
- . . . was threatened in remarks by a student.
- . . . had a weapon pulled on me.

Classroom Disruption

How much of your time in the classroom is directed to coping with disruptive student behavior?
How much does the behavior of some students in your classroom (talking, fighting, etc.) keep you from teaching?

Low Expectations

Of the students you teach, what percentage would you say are low ability?
Of the students you teach, what percentage are behavior problems?

Table B (cont.)

Openness to Student Suggestions

I often change my lesson plans based on student suggestions.
Students should have a lot to say about how the school is run.

Professional Development

How often do you attend professional develoopment courses that are half a day or more in length?
How much training have you had in teaching methods and curriculum content in the last 12 months?
How much training have you had in interpersonal or intergroup relations in the past 12 months?
In some school years, a teacher learns a lot about education, while in other years a teacher doesn't learn much. This year, have you learned much about:
- . . . new materials, new kinds of texts, supplementary materials?
- . . . theories of teaching reading?
- . . . effective methods of maintaining discipline?
- . . . how to handle disruptive students:
- . . . how better to deal with heterogenous classes?

Non-Authoritarian Attitudes

If a pupil uses obscene or profane language in school it should be considered a moral offense. (-)
A few pupils are just young hoodlums and should be treated accordingly. (-)
The threat or use of physical punishment is an effective way of dealing with misbehaving students. (-)

Table C

Item Content of Student School Climate Scales<a>

Community Crime and Gangs

33. Are there any gangs in the neighborhood where you live?
33. Do gang members try to get you to join their gangs?
33. In the last year has either of your parents been robbed on the streets of your neighborhood?

Gangs in School

33. Are there gang members in your school?
33. Do gang members cause a lot of trouble in your school?

Safety

65. Do you usually stay away from any of the following places because someone might hurt or bother you there?
--The shortest way to school (-)
--Any entrances into the school (-)
--Any hallways or stairs in the school (-)
--Parts of the school cafeteria (-)
--Any school restrooms (-)
--Other places inside school building (-)
--Other places on the school grounds (-)
66. In this term in school, have you:
--Had to fight to protect yourself? (-)
--Seen a teacher threatened by a student? (-)
--Seen a teacher hit or attacked by a student? (-)
67. How often do you feel safe while in your school building?
67. How often are you afraid that someone will hurt or bother you at school? (-)
67. How often are you afraid that someone will hurt or bother you on the way to or from school? (-)

<a>Numbers indicate position of the item in the questionnaire.

Table C (cont.)

Disruption

(Note that this scale is partly redundant with the previous one.)
66. (Have you) had to fight to protect yourself?
66. (Have you) seen a teacher threatened by a student?
66. (Have you) seen a teacher hit or attacked by a student?
66. (Have you) been in a class that was totally stopped by a disruptive student?

Individualized Instruction

53. I have a learning plan made just for me.
53. I can work at my own speed in class.

Student Disrespect (Disrespect for Students)

51. Students are treated like children here.
61. Teachers treat students with respect. (-)
61. Teachers do things that make students feel "put down."

Student-Teacher Interaction

58. I talk to some of my teachers about things other than homework.
58. Teachers help me with schoolwork outside of class.

Planning and Action

56. It is hard to change the way things are done in this school.
52. The teachers and principal in this school make plans to solve problems.
52. This school hardly ever tries anything new.

Fairness

51. The school rules are fair.
51. The punishment for breaking the rules is the same no matter who you are.
57. The principal is fair.

Measures

Table C (cont.)

Clarity

51. Everyone knows what the school rules are.
56. The teachers let the students know what they expect of them.
57. The principal runs the school with a firm hand.
56. The principal lets the students know what he or she expects of them.

Student Influence

52. Students have little say in how this school is run. (-)
51. Students can get an unfair rule changed.
52. The student government makes important decisions.
52. Teachers sometimes change their lesson plans because of student suggestions.
52. Students are seldom asked to help solve a problem the school is having. (-)
56. Students have helped to make the school rules.

Grouping

56. Students of different races usually end up in different classes.
56. This school has special classes for slow learners.
56. There are special classes for trouble makers.

Table D

Item-Content of School-Level Scales
Based on Aggregate Teacher Reports

Resources for Instruction

a. This school supplies me with the material and equipment I need for teaching.
b. This school building has the space and physical arrangements needed to conduct the kinds of programs we need.
c. The school's learning program extends to settings beyond the school building for most students.
d. Teachers and students are able to get the instructional materials they need at the time they are needed.

Involvement of Parents and Community

a. How much influence on school policies or practices does a PTO have?
b. How often do...parents help to decide about new school programs?
c. How often do...parents serve as tutors or aides in the classroom?
d. How often (is)...community involvement...sought in developing the school's goals?
e. (How well do) parents and teachers (get along at your school)?
f. Parents and the community are receptive to new ideas.

Teaching Staff Commitment

a. (Are the teaching faculty) involved?
b. (Are the teaching faculty) uncommitted? (-)

Staff Morale (vs. Alienation)

a. Students here don't really care about the school. (-)
b. Our problems in this school are so big that it is unrealistic to expect teachers to make much of a dent in them. (-)
c. I feel my ideas are listened to and used in this school.
d. I want to continue working with the kind of students I have now.
e. (Is the teaching faculty) apathetic? (-)
f. (Is the teaching faculty) cohesive?
g. (Is the teaching faculty) enthusiastic?
h. (Is the teaching faculty) frustrated? (-)
i. (Is the teaching faculty) satisfied?
j. (Is the teaching faculty) tense? (-)
k. (Is the teaching faculty) unappreciated? (-)

Planning and Action

a. How often do you work on a planning committee with other teachers or administrators from your school?
b. The principal encourages experimentation in teaching.
c. Teacher evaluation is used in improving teacher performance.
d. (The principal is) planful.
e. (The principal is) progressive.
f. (The teaching faculty are) conservative. (-)
g. (The teaching faculty are) innovative.
h. (The teaching faculty are) open to change.
i. (The teaching faculty are) traditional. (-)

Table D cont.

Smooth Administration

a. Simple non-time-consuming procedures exist for the acquisition and use of resources.
b. (How well do) teachers and administrators (get along at your school)?
c. Administrators and teachers collaborate toward making the school run effectively.
d. There is little teacher-administrator tension in this school.
e. Our principal is a good representative of our school before the superintendent.
f. The principal is aware of and lets staff members and students know when they have done something particularly well.
g. Teachers or students can arrange to deviate from the prescribed program of the school.
h. Teachers feel free to communicate with the principal.
i. The administration is supportive of teachers.
j. It is hard to change established procedures here. (-)
k. (The principal is) informal.
l. (The principal is) open.

Individualized Instruction and Grading

a. My students mostly work according to individualized learning plans.
b. Students in my classes generally receive grades based on improvement in their performance rather than in comparison with other students.
c. Grades in my classes are typically based on the curve. (-)
d. Grades in this school are typically based on the curve. (-)

School Race Relations

a. (How well do) students of different races (get along at your school)?
b. (How well do) students of different nationalities (get along at your school)?

Interaction with Students

a. In the past two weeks have any students come to you to ask your advice on some problem they were having outside of class?
b. (How often do you engage in) tutoring individual students before or after school?
c. (How often do you engage in) working with students on extracurricular activities?
d. (How often do you engage in) going to games, dances and other student activities?
e. (How often do you engage in) discussing students' personal problems with them?
f. (How well do) teachers and students (get along at your school)?

Integration vs. Segregation by Ability or Conduct

a. Students of mixed ability work together in small groups in my class.
b. Most of my students are assigned to my classes on the basis of their ability. (-)
c. All students in my classroom are of the same general ability level. (-)
d. This school has special classes for slow learners. (-)
e. This school has special classes for high ability students. (-)
f. In this school there are special classes for students who repeatedly misbehave. (-)

Table D cont.

Student Influence

a. I often change my lesson plans based on student suggestions.
b. Teachers and their students work together to make rules governing behavior in the classroom.
c. Students can get an unfair school rule changed.
d. Students help to make the school rules.
e. Students should have a lot to say about how the school is run.

Professional Development

a. How often do you attend professional development courses that are a half day or more in length?
b. (How much in-service training have you had in) teaching methods or curriculum content (in the last 12 months)?
c. (How much in-service...had in) interpersonal or intergroup relations (in the last 12 months)?
d. (Have you learned much about) new materials, new kinds of texts, supplementary materials?
e. (Have you learned much about) theories of teaching reading?
f. (Have you learned much about) effective methods of maintaining discipline?
g. (Have you learned much about) how to handle disruptive students?
h. (Have you learned much about) how better to deal with heterogeneous classes?

Perceptions of Disruption or Lack of Safety

a. (In your opinion, how much of a problem are vandalism, personal attacks and theft) in your school?
b. How much of your time in the classroom is directed to coping with disruptive student behavior?
c. How much does the behavior of some students in your classroom (talking, fighting, etc.) keep you fron teaching?
d. Since school started this year, how many times did you hesitate to confront misbehaving students for fear of your own safety?
e. (How safe is) your classroom while teaching?
f. (How safe are) empty classrooms?
g. (How safe are) hallways and stairs?
h. (How safe is) the cafeteria?
i. (How safe are) the restrooms used by students?
j. (How safe is the) locker room or gym?
k. (How safe is the) parking lot?
l. (How safe is it) elsewhere outside on school grounds?

Table D , cont.

Success Opportunities

a. In this school, students who do well often get special privileges.
b. Any student can earn an A in my class.
c. Some students in my classes earn mostly D's and F's because they cannot keep up with other students. (-)

Use of Grades as a Sanction

a. When a student misbehaves in my class, I sometimes lower his or her grade.
b. (In...how often do you) lower their grades if misconduct is repeated?

Overview of Interim Results for the Alternative Education Program

Denise C. Gottfredson

The School Action Effectiveness Study (SAES) has completed its first year of evaluation activities. A summary or overview of the status of the evaluation and some interim--or formative--evaluation results are provided in this chapter as a quick guide to the 17 action projects. Each of these projects is discussed in more detail in Part 2 of this report, and readers should consult those descriptions for more detail about a specific project.

Conclusive statements are seldom possible at this stage in any large-scale evaluation. Developing, implementing and evaluating social programs takes time. In general, it is too early to reach summative judgments about the effectiveness of these projects, and few will be found in these pages. Information presented here is in the nature of formative evaluation; it should be read in that spirit.

Comments about the nature of the interventions being attempted, the progress made towards their implementation, and the vigor and rigor with which their evaluation is being pursued are, however, timely. In this chapter, as in most of the project descriptions found in Part 2, frank attempts are made to give feedback about the projects. This feedback is intended to provide reinforcement for what appear to be project strengths, as well as impetus to overcome some weaknesses as we see them.

Implementation

The 17 alternative education projects are in varying states of implementation. Some are still floundering with start-up efforts and with unclear goals and others are extending or replicating efforts they have tried elsewhere or at an earlier time. A brief characterization of each project is presented in Table 1. See the individual project descriptions for more comprehensive accounts.

Some common themes arise from a reading of the more detailed project descriptions in Part 2 of this report. First, a difficulty experienced by many of the projects relates to the

---

I am grateful for the comments on a draft of this chapter by Phyllis Betz, Vic Cooper, Deborah Daniels, Gary D. Gottfredson, Joe Nathan, Jane St. John, Dave Reiss, and Sally Wisotzkey. Opinions expressed are, however, solely my own. This report covers, for the most part, the period ending August, 1981.

timing of the grant award notification. For most projects, notification came just at the beginning of the school year, leaving them in a state of uncertainty. Many of the projects thus entered the 1980-81 school year without plans, and unable to begin work immediately, but feeling pressure to do so. The long lead times required for administrative decision making (e.g., in the Chicago Board of Education project), or the inflexibility of school system arrangements, often thwarted implementation further.

In a few cases, grant awards were made in winter, between semesters. Again, this appears to be an awkward time to begin a school-based project. Presumably notification of grant awards would be best made near the close of the academic year preceding project start-up, if not earlier, to facilitate project planning and smooth implementation.

Second, school system changes, such as grade structure reorganization, a change of administrators, or reductions in staff size, create serious problems for project implementers. Sometimes, as in the Kalamazoo project, interventions were planned in collaboration with personnel who were no longer there when the project began operation.

Third, ambiguity about project staffing or difficulties with staff turnover can impede implementation. For example, the designation of project directors was delayed in both the Chicago Board of Education and St. Paul projects. In the case of the Chicago Board project, ponderous administrative machinery seems to be the cause of this delay. The Bronx, Hayward, Houston, and Harlem project directors were replaced several months into the project, and other projects have experienced instability in their staffing patterns. These problems naturally cause difficulties in implementation.

Fourth, projects differ in the clarity of their implementation plans. Projects without clear plans, and projects that do not engage in systematic planning, must struggle harder for success.

Finally, evaluation taxes the resources and patience of most projects. Ambiguities in the RFP regarding the nature of the evaluation and the level of resources projects would have to allocate to evaluation activities left projects unprepared for, and sometimes bitter about, the intensive activities they were expected to undertake. Specifically, few projects expected evaluation considerations to influence the selection of students for their interventions. Establishing and maintaining a rigorous evaluation design without a prior agreement among all affected actors requires intensive negotiation and careful monitoring. In addition, some projects

were unprepared to provide individual-level data even for youths receiving project services, and others had difficulty providing data for comparison students. Furthermore, the SAES student survey was also difficult to implement, not only because of the sensitive nature of many questions, but also because its administration required a large commitment of personnel resources, because some projects expected to conduct a different kind of evaluation, and because of inadequate or unrealistic budgetary planning for evaluation. The response rates shown in Table 2 are in part an indicator of the level of resources projects were willing to allocate to this important evaluation activity.

These observations have five implications for the future efforts of these 17 projects, and for similar projects attempted in the future:

1. Notification of funding should be made before the end of the academic year preceding the anticipated activity.

2. Projects in school systems undergoing other major administrative or structural changes should be avoided.

3. Staffing plans should be well considered and made in advance of project start-up.

4. Project implementers should be proactive in the development of plans. The importance of clear plans can scarcely be overstated.

5. Evaluation requirements and guidelines for the level of resources to be allocated to the evaluation should be specified in the RFP; evidence of the project's intent to and ability to comply with the guidelines should be required. Projects should be site-visited prior to funding, and detailed written agreements regarding access to information and experimental designs should be formulated at that time. The content of these agreements should form one basis for making decisions about funding.

Overview of Evaluation Results for the Entire Program: Outcome Evaluation

Only rarely were the conditions necessary for making inferences about project effectiveness (see Chapter 4) met during the first program year. The evaluation designs for the first year are summarized in Table 3. Some projects used the first year as a planning and staff-training period and hence did not start providing services. Some provided services, but only on a trial basis. Others provided services for the better part of the year but had already selected the students by the time the national evaluation staff contacted them to set up selection procedures that would result in a defensible

experimental or quasi-experimental design. The evaluation was funded and began work on 29 September 1981. A longer start-up period for the evaluation, with the opportunity to work with projects before they began implementation, would have been desirable.

Projects Implementing True Experiments

Two projects attempted experiments during the Spring semester. Peer Culture Development (PCD) and the Chicago Board of Education's Project RETAIN randomly assigned students to treatment and control groups. Although both projects met with some obstacles as they implemented their designs, the experience was valuable because they were able to plan strategies to overcome these obstacles for the following semester.

Peer Culture Development. PCD developed a large enough pool of referrals who volunteered for PCD classes to randomly assign the students to treatment and control groups. But for a few class periods, all students in the pool had to be placed in the treatment group. And in some classes, student assignments were made non-randomly, persons who had been designated as "treatment" students did not participate in the PCD classes, and some control students received PCD services. The effective "N" for the experiment was reduced considerably for these reasons, and data for certain classes had to be excluded from the study, weakening the experimental design. For statistical purposes persons were treated as if they had experienced the experimental condition to which they were assigned in the randomization process. The results of the experiment apply only to those classrooms included in the experiment. In addition to these limitations, PCD discovered that random assignment altered the composition of their classes because the group leaders had less control over the designation of participants. The project personnel, on the basis of their first semester experience, planned and implemented procedures for the following semester that would increase the pool of eligibles and ensure a more desirable mix of students in the group.

Despite these limitations, useful information, described in the project summary by St. John (this volume), was obtained from this experiment. The PCD treatment as implemented reduced disciplinary infractions, and had positive effects on interpersonal competency and practical knowledge; the treatment had an unanticipated negative side-effect on belief in conventional rules. This project's exemplary evaluation is continuing and is being strengthened in the 1981-82 school year.

Chicago Board of Education. CBE had little trouble obtaining ample referrals for its program, but the randomization process

broke down in some of its schools, resulting in non-equivalent comparison groups in those schools. Thus, although post randomization checks did not detect failures of randomization in most CBE schools, the way randomization was conducted on-site weakened the design.

Additional problems involved the questionnaire administration, an important design consideration because even a true experiment is diminished in value without sound measurement of the outcomes of interest. Low response rates for the students in the study resulted in small N's and biased samples. To complicate matters further, CBE censored most self-report delinquency items in the questionnaire. Finally, only two months elapsed between the date the students began receiving RETAIN services and the date the questionnaire was administered. These conditions worked against detecting treatment-control differences, especially on the survey measures.

A decision by the Deputy Superintendent for Field Services of the Chicago schools to delete most self-report delinquency items from the student questionnaire created a large problem for the evaluation of both Chicago projects. PCD was able to obtain delinquency data from an alternative source (police records), but CBE refused to do so.

These obstacles notwithstanding, the two Chicago projects were both partially successful at implementing experimental designs.

Other Projects

Inferences about the effectiveness of the remaining projects on any outcome measures cannot yet be made. We attempted to evaluate the effectiveness of the seven projects that identified non-equivalent comparison groups by statistically adjusting for differences between the treatment and control groups that existed before the treatment began. This activity proved to be of little value because we did not have pre-test measures of the most relevant characteristics to use as statistical controls. Such adjustments are expected to be more useful as a fall-back design next year when good pre-test measures will be available in most cases, at least for many of the youths.

Although any of a number of projects could be used to illustrate the problem, the Constitutional Rights Foundation design provides a good example of the flaws inherent in this type of analysis. This project involves two treatments--a Youth Committee and an Options Class. The selection criteria for the two treatments differed. The Youth Committee members were selected from a pool of students who had exhibited leadership; the Options Class members from a pool of

students who had experienced academic difficulty. One would expect, given the selection procedures, that the Youth Committee participants would be considerably more interpersonally skilled than the average student in the school. The the Options Class participants would be expected to be lower academically than the typical student in the school. This pattern of differences showed up when we compared the means on outcome measures for the two treatment groups with each other and with the mean scores for a random sample of other students in the school (see raw means, Table 4). Only weak proxy measures of leadership ability, prior academic difficulty, and other group selection criteria were available, so we were unable to adjust the observed differences in the outcomes for the pre-existing differences. The adjusted "change in outcome" columns in Table 4 reflect the difference between youth committee participants and the students in the random sample and between options class students and random sample students that remains after preexisting differences on family background measures, such as parents' education level, are taken into consideration. The adjusted mean scores for the groups still show what can most plausibly be interpreted as the pre-existing pattern of differences: The Youth Committee is highest, the random sample is in the middle, and the Options Class is at the bottom of the distribution on most measures of social development. The group differences on outcome measures cannot be interpreted as indicators of program effectiveness (or ineffectiveness, in the case of the Options Class) because the hypothesis that the differences reflect pre-existing differences has not been laid to rest. In the CRF case, we have additional evidence that the differences may not have resulted from the program: The Youth Committee intervention was not implemented as fully as was anticipated last year, and it is primarily intended to create structural rather than individual change.

Summative Evaluation Designs for the Second Project Year

The designs for making inferences about project effectiveness (summative evaluation) for most projects will be strengthened considerably in the second year of operation. An overview of those designs is given in Table 5. Some designs are more difficult to implement than others, and some project settings are less amenable to experimental activities. The level of resources committed to evaluation activities has in the past varied across projects; projects may be expected to show disparate levels of effort. These issues all influence the likelihood that planned evaluation designs will be implemented. Such concerns contribute to the overall assessment of evaluatability shown in the far right column of Table 5.

Three projects--Bronx, Charleston, and Peer Culture Development--stand out as highly evaluatable projects. Three projects--Harlem, New Jersey, and St. Paul--are low in evaluatability, at least in terms of summative evaluation. The remaining 11 projects are evaluatable, but their designs are weak in some way, or there is uncertainty that the design will be implemented as planned.

Summative evaluation is, of course, not the only concern of an evaluation. In some cases where the designs to assess project effectiveness are rather weak, or even dismal, other "ways of knowing" (Tharp & Gallimore, n.d.) can provide leads for more careful examination in a replication elsewhere or at another time. Because the time period for the Alternative Education Program is limited, however, the possibilities of pursuing these leads are also likely to be limited. This makes summative evaluation of these action projects a very important evaluation concern.

Project Descriptions

Project Size and Characteristics of Students Served

Tables 6 through 8 show the number of students served by each project and the ethnicity and socioeconomic level of these students as of April 1981 where this information is available from the SAES student questionnaire data collected in late Spring, 1981. School total enrollments are also shown. Some projects (e.g., the Chicago Board of Education's project) aim to provide intensive services to a small number of students. Others (e.g., the Kalamazoo project) provide only indirect services to specific students because they are primarily aimed at school structural changes.

The ethnicity of students served varies across projects. Some projects (e.g., Plymouth) serve predominantly white studentries.<1> Others (e.g., Harlem, Compton, and Charleston) serve predominantly black populations. Houston and Puerto Rico serve Spanish-speaking or Spanish-surnamed populations, and the Bronx project serves a mixed Hispanic and black population. The Hayward project serves an American Indian population.

The level of parents' education is a measure of the relative social status of the clientele being served. Table 8 shows that there is a considerable range of parental education across the projects. (The table entries show average parental education, (mother + father)/2, where "6" means completed four years of college, and "3" means finished secondary school.) In Puerto Rico, for example, the average student's parents have not completed high school, whereas in Pasadena the average student's parents have graduated from college.

Students receiving direct services in the Alternative Education Program are remarkably similar in ethnicity and family educational background to other students in the same schools. Isolation of minority or low socioeconomic status students does not generally appear to be occurring.

School and Community Context

Tables 9 through 11 describe the community and school contexts, student composition, and the characteristic school governance for the 58 schools from the 15 projects included in the first annual survey administration (Miami and Milwaukee were not included). All characteristics reported in these tables are measured by scales constructed from items in the student questionnaire. See Chapter 5 of this report for descriptions of the scales.

These tables are abstracted from more detailed profiles based on both teacher and student surveys, which were distributed to project administrators at the end of the first program year. The characteristics presented here are based on student survey scales only. They are intended to provide a characterization of gross project-to-project differences in the communities and students served, as well as a description of the school environments.

Each row in the tables summarizes a school's placement in the distribution of mean scores for all 60 schools. A '+' indicates that the average score of all student reports in the school was above the 75th percentile for all schools. A '-' indicates that the average fell below the 25th percentile. A '*' indicates that items used in the scale were not measured at the school.

The tables indicate that the projects with alternative schools that are separate from the school system--Houston, Compton, and Lac Courte Oreilles--have the most positive climates. Students from all three of these projects report that the rules in their schools are fair. Students in Compton and LCO report high levels of student-teacher interaction, planning, and clarity of rules. Students in Houston and Compton are less alienated and more attached to school, and students in Houston and LCO report lower levels of school disruption than do students in most other schools. These three projects are among the four with the highest levels of delinquency.

Another general pattern which emerges from Tables 9 through 11 is that the Puerto Rico project operates in an exceptionally pleasant climate: Low levels of disruption, delinquency, and alienation and high levels of attachment, belief, and each of the five governance

characteristics are reported. Charleston and Kalamazoo also operate in fairly positive settings, although not nearly as positive as Puerto Rico.

The Chicago projects, on the other hand, are operating in less than ideal conditions. One or the other of the Chicago projects is extremely low on four of the five governance characteristics, and students in most of the RETAIN schools report high levels of community crime and alienation. Interestingly, the PCD schools report low levels of disruption. The New Jersey and Virgin Islands projects also have extreme negative scores on many dimensions, although Virgin Islands, despite high community crime and low attachment and belief, scores at the low end of the delinquency distribution.

Finally, it is interesting that students in both New York City projects (PREP and Jazzmobile) are among the three lowest-scoring projects on belief in the validity of conventional rules, and that students in projects located in the more affluent Kalamazoo score extremely low on attachment to school.

Program Elements

The program models for the 17 alternative education projects are very different. Although all projects incorporate at least one of the desired project models specified in the RFP, the degree of overlap between the solicitation requirements and the actual program models varies considerably. The Charleston project, PATHE, includes almost every program element described in the RFP. The Peer Culture Development project, on the other hand, focuses most of its resources on only one of the nine program models--altering peer group experiences.

In short, the Alternative Education Program includes a broad range of projects. Some are attempting to implement a wide variety of interventions, and others, only a few kinds. Projects implementing only a few interventions appear, in general, to be focusing most of their resources on those interventions and may therefore be implementing them with considerable strength. By doing so, however, they may be neglecting many additional opportunities for intervention. Those projects implementing a variety of kinds of intervention are hitting the system in many places at once, but run the risk of diluting the strength of specific interventions by overextending themselves.

Four general types of planned intervention categories may be identified, and the following subsections illustrate interventions in each category.

Method of Instruction. One type of intervention aims to alter the content or method of instruction in the schools in an attempt to

improve youths' motivation and commitment to education as well as their chances of success in school. Two main strategies are used: Projects attempt to individualize instruction either through tutoring by teachers, peers, or outside agents, or through the development of individualized learning plans. The Charleston, Chicago Board of Education, and Compton projects use both strategies. Kalamazoo uses tutoring, and the Bronx project uses individualized learning plans.

Another strategy for altering the delivery of academic instruction is curriculum development. Some projects focus on altering the content of the curriculum. Jazzmobile, for example, offers courses in the arts, and Project STATUS in Pasadena offers Options Classes focusing on English and social studies, and has initiated a Leadership Training Class. Compton and the Jewish Vocational Services project in Milwaukee emphasize practical vocational skills development in their courses. The Bronx and Houston projects offer instruction in cultural heritage. Curriculum development also comes in the form of methods development. Project STATUS uses small groups and simulations in its classrooms; Charleston uses Student Team Learning; and the Chicago Board of Education, Lac Courte Oreilles (Hayward), and Compton use computer-assisted instruction.

Reward systems are the focus of some projects' interventions. The Miami project model calls for the establishment of a token economy system. Project PREP in the Bronx tries to implement a "time-out" room for students who are falling behind academically or disrupting the classroom. Non-traditional rewards--such as field trips, sweatshirts, opportunities to display artwork in the community--are used in the Bronx, the Virgin Islands, Pasadena, Harlem, and Houston.

Finally, the St. Paul project is implementing a structure to alter the nature or process of learning. It uses Action Learning Projects wherein students design and execute projects to help solve a school or community problem. Teachers serve as managers of learning rather than instructors.

Service delivery. A second major program component focuses on improving the schools' delivery of services. This is often accomplished through counseling: Peer Culture Development, St. Paul and Charleston use peer counseling, and Plymouth and the Bronx use individual counseling. Milwaukee, Puerto Rico, Houston, and Miami focus on vocational counseling. In Pasadena, a project counselor with a caseload of about 120 students provides services to both students and parents.

Various classroom management techniques are also used in attempts to meet the needs of students: Virgin Islands and Kalamazoo implement some kind of a "family group" period. During this period, small groups of students and a teacher discuss problems the students are experiencing. In Miami, classrooms are managed using principles of applied behavioral analysis, and in Compton and the Virgin Islands teachers are selected and trained to avoid negative reinforcement. In Pasadena, students in the various project components formulate their own rules for classroom management.

Student participation in decision making is another affective strategy. Charleston, Kalamazoo, and Project STATUS involve student groups in school-level decision making. The Puerto Rico project also uses student involvement, but the focus is on involving students in making decisions about their own education. A number of projects attempt to increase students' involvement in extra-curricular activities either by starting up new activities and clubs (Puerto Rico, Charleston), by providing opportunities for students to go on trips (Charleston, Bronx), or by organizing sports teams (Houston).

School organization. A different strategy aimed at meeting students' affective needs might be called general school climate improvement. For example, the Charleston project focuses on climate improvement by organizing a "school pride campaign" aimed at improving teacher, student and community perceptions of the school through the use of the media, pep rallies, clean-up campaigns, etc. The project also sponsors faculty team-building activities to improve morale. Or, for example, the Pasadena project aims to influence school climate through its Youth Committee.

The St. Paul project, like the Charleston and Virgin Island projects, focuses heavily on public relations activities. It has been active in helping schools explain to their constituencies what it is trying to do. It has worked with school staffs to get students involved in writing for community newspapers an providing pictures for those papers. A group of students has organized a memorial fund for an outstanding teacher, and shared in publicity efforts.

Altering overall school organization and management is another school organization intervention. The Kalamazoo project is exemplary in this area. Its program model calls for establishing and maintaining an organizational structure composed of area task forces, a building steering committee, and an advisory council that engages students, teachers, building administrators, district-level school administrators, parents, and juvenile jus-

tice service administrators in joint decision making and in the implementation of all aspects of the project.

Charleston is a project which alters the management of the school in more specific ways: A disciplinary referral procedure aimed at increasing the consistency of rule enforcement is established, and business-education partnerships between community business representatives and building principals are formed. These partnerships offer expert management assistance to school administrators. Other projects, including the Puerto Rico and St. Paul Projects, also are working on developing partnerships with community businesses.

The St. Paul project aims to alter school organization and the way the schools are perceived by the public by increasing school-community linkages. It also seeks to improve school governance by providing training for teachers, working with administrators to foster improvements in school rules and their administration, and getting students involved in working with teachers on the development of these rules. In addition, it is developing an advisor-advisee system in its schools. This system is intended to provide small groups for students to discuss problems and obtain information, and to receive more individual attention than would otherwise be possible. It is also working to create more active student councils and has created a school advisory council with parents, students and staff represented.

Community involvement. The fourth major program component is community involvement. Community involvement is encouraged both for the purpose of channeling community resources into the schools and for helping individual students or clients. Parents are the primary community contacts. Charleston, for example, involves parents in school improvement activities. Other projects (e.g., Compton, Chicago Board of Education, Puerto Rico, Bronx) seek to increase parental involvement in their own children's educations by informing them of students' progress and of school activities. And, the Pasadena project makes use of an Advisory Action Committee and the parents of students participating in its project components as resources for field trips and internships. Finally, business contacts are used as resources to improve schools (e.g., St. Paul) and to employ youths (Milwaukee).

The foregoing classification is intended to provide the reader with a general overview of the types of programs funded under the alternative education initiative. It is based for the most part on project plans. Detailed descriptions of the program plans and of what was actually implemented at each project site during the first year of operation are

found in the descriptions of individual projects.

Organizational Focus and Institutional Change

OJJDP stressed an organizational focus and an organizational change approach in its request for proposals for the initiative, and repeated these themes continually during the first program year. The RFP called for programs that seek to reduce delinquency by "changing the structure and the educational processes of schools" in such a way that students would maintain or develop a stake in academic achievement and conformity to conventional rules, and would be less likely to become alienated and engage in delinquent activities.

One can imagine alternative organizations of the American educational enterprise. Compulsory education as it exists today might be abolished in favor of work-study relationships between adolescents and an employer. Schools might exist solely as resource centers. Or, they could supplement a traditional curriculum with on-the-job training components. Large schools could be broken up into clusters--schools within schools. Less drastic organizational changes might include altering grading practices so that students do not experience academic failure in school, or giving student coalitions the responsibility for developing and maintaining school discipline codes. The possibilities are nearly endless.

Considering this wide range of possibilities, the projects funded in the alternative education initiative focus minimally on organizational change. Instead, the projects generally stress an alternative approach to dealing with a troublesome subpopulation of youths. Some do, in addition, direct their efforts toward influencing specific school policies and procedures, and some attempt to augment the existing school curricula with materials or methods having wider appeal. Some create organizations outside the public school system but provide schooling that resembles traditional education in a modified setting and with somewhat different emphases. But none of the projects concentrates on making changes in the basic educational structure of the kind that would radically change the condition of youth in contemporary society (cf. Gottfredson, 1981; Greenberg, 1977; President's Science Advisory Committee on Youth, 1967).

The projects might be ranked on criteria related to organizational change focus, such as the percentage of resources expended on attempts to change existing structures, practices, or policies, or the percentage of critical benchmarks met related to organizational change. Such a ranking would be misleading because it fails to consider the history of the organiza-

tions and the organizational settings in which the projects operate. The Charleston, Kalamazoo, and St. Paul projects--all school system projects--would rate high, while the Houston, Harlem, and Miami projects--all community-based organization projects--would rate low on a measure of staff time devoted to changing existing public school practices. This ranking ignores the prior work done by the CBO's in developing an alternative or addition to the existing system. The proof of the pudding might be in the degree to which the various alternatives are adopted by or replace the system to which they provide an alternative.

Systems are not easily changed. The environment must be ready to accept change; i.e., there must be dissatisfaction with the status quo. An alternative model for organization and management of the system must be present, as well as a planned process for managing the change. Charismatic and dynamic leadership also facilitates change. Finally, the cost of the proposed change must be less than or equal to the cost of maintaining the present system. These criteria were incorporated into a list of indicators of the likelihood of school change, by Noel Day of Polaris (the Technical Assistance Contractor for this initiative). Table 12 is adapted from the criteria used in Day's rating, and Figure 1 presents the ranking. Day did not rate several projects (Charleston, Puerto Rico, Miami, New Jersey) with which he was unfamiliar. We have placed these projects into the figure where they appeared to belong, using interpolation based on information developed by our evaluation. On the whole, the assessments made by Day jibe well with what we know about the projects. We adjusted Day's rankings for three projects--Compton, Pasadena, and Virgin Islands--on the basis of new information about these projects.

These rankings, of course, are an interpretation of hunches about the likelihood of creating change in existing school systems. Although we understand OJJDP's desire to create such change, and to sponsor demonstration projects likely to create it, two considerations, make us doubt the utility of an excessive focus on institutional change in school systems at present.

First, one widely held, legitimate perspective on the public schools is that the public schools would require such extensive restructuring to serve all students well that, rather than seeking to change them, we should be seeking alternatives to them. It has not escaped our notice that those projects to the far left (no pun intended) in Figure 1 are creating alternatives to the public schools.

Second, it is at present premature to assess the desirability of institutionalizing any of these projects. They have not yet demonstrated effectiveness at reaching their stated goals and objectives. Despite the apparent enthusiasm on the part of some education departments for replicating or extending action project activities elsewhere, these projects are of uncertain effectiveness. We are convinced that assessing project effectiveness is at present more important than focusing on extension or replication. Widespread public perceptions of the ineffectiveness of social programs are probably due, at least in part, to premature efforts to implement new programs everywhere at once without thorough and systematic testing and development.

References

Gottfredson, G. D. Schooling and delinquency. In S. E. Martin, L. B. Sechrest, & R. Redner (Eds.), New directions in the rehabilitation of criminal offenders. Washington, D.C.: National Academy Press, 1981.

Greenberg, D. F. Delinquency and the age structure of society. Contemporary Crisis, 1977, 1, 189-223.

President's Science Advisory Committee on Youth. Youth: Transition to adulthood. Washington, D.C.: U.S. Government Printing Office, 1967.

St. John, J. Peer Culture Development (PCD), Chicago. This volume.

Tharp, R. G., & Gallimore, R. The ecology of program research and development: A model of evaluation succession. Unpublished manuscript. (Available from Roland Tharp, Department of Psychology, University of Hawaii, Honolulu 96822, or from Ronald Gallimore, Department of Psychiatry and Biobehavioral Sciences, UCLA, Westwood, California 90024. (A secondary account of the evaluation succession approach is available in I. McNett, KEEP early education project . . . from research to evaluation to success, APA Monitor, December, 1981, pp. 8-9, 33.)

Footnotes

1. These tables may not, however, convey the entire picture. For example, according to the St. Paul quarterly report to OJJDP for the fourth quarter of 1982, 36% of the cummulative number of students receiving direct services from the project were Black. This contrasts with the proportion of Black students in the schools estimated from the student survey. Unfortunately, these percentages can not be estimated dependably using questionnaire data for this project. A number of conditions, including differential mobility of students of different ethnic groups and differential non-response for different ethnic groups, sometimes makes the racial composition of a school's studentry difficult to estimate from survey data.

Figure 1

Ranking of Alternative Education Projects on Likelihood of Institutionalization in School System

			RETAIN--Chicago
JVS--Milwaukee	Puerto Rico	CRF--Pasadena	Plymouth
Jazzmobile--Harlem		Miami	PREP--Bronx
Hayward	Compton	PCD--Chicago	St. Paul
Houston	New Jersey	Kalamazoo	Virgin Islands
			PATHE--Charleston

LESS LIKELY ---------> MORE LIKELY

Note. Adapted from a figure by Day (1981). The Charleston, Puerto Rico, Miami, and New Jersey projects were omitted from Mr. Day's ranking. The evaluation staff produced rankings for these projects and modified Day's rankings for three other projects--Compton, Pasadena, and Virgin Islands--based on additional information.

Table 1

Overview of Implementation Status of the Alternative Education Projects: August 1981

Project	Implementation Status
Compton	Is providing some of the anticipated services to students. Has had high staff and student turnover and difficulty implementing some project components.
CRF, Pasadena	Is implementing some of its anticipated project components, but not others; used the Sprint as start-up time to write experimental curriculum. Has had difficulties attributable in part to a lack of careful proactive project planning, and in part to limited school system commitment to the project.
PCD, Chicago	Is replicating some well practiced procedures in nearly the form intended. Its range of interventions is fairly narrow, but they appear to be implemented with considerable strength.
CBE, Chicago	Is providing some of the intended services, although not with the strength or integrity initially hoped for. Experienced difficulties starting up, and has not been characterized by systematic proactive planning.
Kalamazoo	Is implementing part of the activities intended. Experienced difficulties due in part to changes in school administration not foreseen when the project was planned. In addition, the project director does not have control over sufficient personnel resources to implement all components of the program fully.
Bronx	Is providing many of the anticipated services to students, although the strength of these is unknown at present. Experienced turnover at the project-administration level.
Jazzmobile, Harlem	Is providing arts instruction, although the exact nature of these activities remains unclear.

Table 1 (cont.)

Project	Implementation Status
Charleston	Is implementing a wide range of interventions; some are not implemented in the form or with the strength intended. May be overextending itself in attempts to do too much.
Houston	Is continuing the relatively smooth operation of an alternative school already in existence, but has not extended activities much.
Virgin Islands	Is implementing one of its major interventions, but has delayed the implementation of others.
LCO, Hayward	Has installed computer terminals, but is implementing little beyond this limited intervention.
Puerto Rico	Is implementing most planned interventions, and is changing some components to strengthen them.
Miami*	Implemented a six-week summer pilot test.
Plymouth*	Implemented some direct services to students.
New Jersey*	Did not begin project implementation in the past year.
Milwaukee*	Began implementation of direct services near the end of the past year.
St. Paul*	Implemented several project components. The acting project director's appointment as project director was delayed, slowing implementation.

*These projects were funded four to five months after the others.

Table 2

Student Questionnaire Response Rates: Spring 1981

Project Name	School Code	N in Sample	Response Rate
Plymouth	31	84	.87
	41	85	.87
	42	111	.66
	43	94	.59
Kalamazoo	318	298	.76
	327	278	.80
Charleston	242	323	.96
	741	350	.95
	742	312	.95
	743	338	.83
	751	308	.84
	754	333	.84
	755	338	.89
	944	310	.96
	951	386	.59
Compton	101	92	.39-1.00<b>
Jazzmobile	88	819	.33
Virgin Islands	1101	295	.88
Lac Courte Oreilles	1201	125	.55
Chicago	6180	233	.86
	5880	247	.88
	5750	263	.87
	5090	161	.14
	1240	376	.46
	1340	367	.59
	1430<a>	---	---
	2300	163	.91
	4440	131	.96
	4550	236	.49

a. See Peer Culture Development

b. Information from different project staff members about the school enrollment at the time of the questionnaire does not converge.

Table 2 (cont.)

Project Name	School Code	N in Sample	Response Rate
Constitutional Rights Foundation	70	291	.79
	82	324	.58
Peer Culture Development	1370	334	.65
	1430	356	.59
	1820	360	.79
St. Paul	210	306	.37
	212	300	.59
	230	302	.66
	342	324	.89
	352	302	.56
Puerto Rico	801	556	.90
	802	230	.83
	803	596	.67
Bronx	22	296	.70
	55	183	.81
	63	152	.64
	64	189	.79
	82	309	.74
	117	333	.71
	132	156	.69
	145	301	.34
	147	311	.39
	148	297	.61
	166	264	.81
	229	355	.77
New Jersey	001	820	.59<c>
	002	299	.69
	003	375	.40
	004	379	.85
Houston	1001	84	.97

---

c. Response rate for this school estimated (no roster available).

Table 3

Summary of Experimental Designs
for First Program Year

Project	Design allows for comparison of program and non-program:		
	Students		Schools
	equivalent	nonequivalent	(all nonequivalent)
Compton	no	no	no
CRF, Pasadena	no	yes	no
PCD, Chicago	yes	yes	yes
CBE, Chicago	yes[c]	yes	yes
Kalamazoo	--[b]	--[b]	yes
Bronx	no	yes	no
Jazzmobile, Harlem	no	yes	no
Charleston	no	yes	yes
Houston	no	no	no
Virgin Islands	no	yes	no
LCO, Heyward	no	yes	no
Miami	--[a]	--[a]	no
Plymouth	no	yes	no
New Jersey	--[a]	--[a]	no
Milwaukee	--[a]	--[a]	no
St. Paul	--[a]	--[a]	no
Puerto Rico	no	yes	no

a) Direct services to students did not begin or began on a partial basis during the first program year.

b) This project, by design, did not offer direct services to youths during the first program year.

c) Design deteriorated on implementation.

Table 4

Details of Analysis of Effects with Protected
Tests of Significance for Project STATUS

Criterion Variable	Raw Mean: Youth Cmte.	Raw Mean: Options	Raw Mean: All Others	Adjusted change in outcome due to: Y. Cmte.	Adjusted change in outcome due to: Options
Muir					
Self-Reported grades	3.12	2.56	2.65	+.42*	-.50
Involvement	2.05	.01	-.38	+2.24*	a
Elliot					
School nonattendance	2.74	1.81	1.64	+.81*	a
Practical knowledge	13.01	11.56	12.47	+.58	-.90

[a]Since the Options Group was correlated less than 0.10 w/ the criterion, it was not included in the analysis.

*Significantly different from adjusted mean for all others, $p < .05$.

**Significantly different from the mean for all others, $p < .01$.

Table 5

Evaluatability Summary
for Second Program Year

Project	Brief Description	Strength of Design	Likely Integrity of Design	Overall Evaluatability Rating
Compton	Random assignment of new referrals to program and control. Most program slots are filled by returning students. Pre-treatment school and police records for both groups. Extremely small sample size works against detecting differences.	4-	1	+
CRF, Pasadena	Non-equivalent comparison group composed of volunteers, referrals, etc., who did not enter program. SAES pretest on all.	2-	2	+
PCD, Chicago	Random assignment to treatment and control for all students. Large sample size. Court and self-reported delinquency data available.	3+	2	++
CBE, Chicago	Random assignment to treatment and control for all students. Randomization process not carefully monitored.	3-	1	+
Kalamazoo	Most change is expected to occur at the school level rather than to individuals being served. The design will allow for comparison of change from year one to year two of the program school with a control school using the SAES questionnaire. Comparison of students directly served and not served will be weak.	2	2	+
Bronx	Random assignment to treatment and control. SAES pretest on both groups.	4	2	++
Jazzmobile, Harlem	Project has no control over assignment, but school's assignment is thought to result in nearly equivalent groups of treatment and nontreatment youths. Likelihood of measuring outcomes is extremely low.	1	0	-

CONTINUED

2 OF 3

Table 5 Continued

Project	Brief Description	Strength of Design	Likely Integrity of Design	Overall Evaluatability Rating
Charleston	Random assignment to treatment and control in all seven schools to assess individual-level, and two comparison schools to assess organizational level, program effectiveness. Delinquency measures limited to official records in middle schools, but pre and post treatment official records available. Large sample size.	3+	2	++
Houston	Design allows for comparison of students in two program components (nonequivalent groups) and a small nonequivalent no-treatment group. Pretreatment data from records available for all groups.	2	2	+
Virgin Islands	Regression discontinuity design for one project component (non-equivalent groups with SAES pre-test as a backup). Small sample size.	2	1	+
LCO, Hayward	Evaluation of PLATO component only. Nonequivalent groups with SAES pretest on relevant outcomes.	2-	2	+
Miami	Comparisons of nonequivalent treatment and control groups. Pre-treatment data from records for both groups, and pre-treatment skill test data for both groups.	2	2	+
Plymouth	Random assignment to program and control with SAES pretest for the two high school components. Nonequivalent comparison groups with SAES pretest for middle school component. No comparison group for out-of-school (Growthworks) component because program serves all problem youths. Growthworks component not evaluatable.	2	2	+

Table 5 Continued

Project	Brief Description	Strength of Design	Likely Integrity of Design	Overall Evaluatability Rating
New Jersey	Non-equivalent comparison groups of students served and not served. Pre-treatment data from records.	2	0	-
Milwaukee	Random assignment to treatment and control groups with SAES pretest for both broups. Multiple follow-up using SAES survey and data from official records.	4	1	+
St. Paul	Most change is expected to occur at the school level rather than at the individual level. Comparison of students receiving and not receiving service will be extremely weak. Design allows for comparison of program and nonprogram schools on a non-SAES questionnaire. No relevant comparisons on any measure of delinquency. Schools undergoing concurrent administrative reorganization.	1 +	2	-
Puerto Rico	No experimental control over program access; but small preexisting differences between service recipients and other students, large-sample multi-wave questionnaire and official delinquency data allow for a non-equivalent control group design.	2+	2	+

Note: The rating schemes used for strength of the designs are as follows: random assignment of subjects to treatment and control conditions with pre-treatment measures on the relevant outcomes is given a code of "4." Random assignment without pre-treatment data gets a code of "3." Identification of a nonequivalent comparison group (or school) with pretreatment information is coded as "2," and nonequivalent comparison groups without pretreatment data is coded as "1." Pluses appended to these codes indicate strong points (such as large sample sizes or multi-level designs) and minuses indicate weaknesses such as extremely small sample sizes, incomplete or non-existent data on delinquency, or flawed pretesting conditions.

Integrity of design is a three-category forecast. A "2" means that faithful implementation is highly likely; a "1" means that faithful implementation is problematical, and a "0" means that faithful implementation appears unlikely.

Table 6
Total Enrollment and Number of Students Receiving Direct Services
as of April, 1981

Project[1]	School	Total Enrollment[2]	Number of students Served[3]
Compton	Alternative School	61	92[a]
Constitutional Rights Foundation	Muir S. H.	2120	111
	Elliot J. H.	1325	121
Peer Culture Development	Lakeview	1366	89
	Curie	3065	87
	Harrison	1100	82
	Spry	--[d]	26
	Pope	--[d]	32
	Edwards	--[d]	16
	Hearst	--[d]	28
	Nettlehorst	--[d]	29
	Ravenswood	--[d]	34
Chicago Board of Education	Lemoyne	515	14
	Bontemps	756	15
	Blaine	718	15
	Gage Park	1400	15
	Bowen High	2700	14
	Lakeview High	1366	14
	Thorpe	933[b]	15
	Nightingale	736[b]	15
	Sheridan	2114[b]	15
Kalamazoo	Milwood	657	0
Bronx	CJHS #22	1003[b]	9
	CES 55	717[b]	15
	CES 63	757[b]	15
	CES 64	1065[b]	16
	CHHS 82	824[b]	17
	CJHS 117	1071[b]	14
	CES 132	673[b]	18
	CHHS 145	1119[b]	6
	CIS 147	1499[b]	18
	CIS 148	861[b]	17
	CIS 166	225[b]	9
	CIS 229	589[b]	1

Project[1]	School	Total Entollment[2]	Number of Students Served[3]
Jazzmobile	Intermediate School	506	364
	Elementary School "A"	504	146
	Elementary School "B"	446	134
Puerto Rico	Santiago Gonzalez	554	82
	Ruiz Belvis	632	39
	Dr. Alfredo M. Aguago	1059	120
Charleston	C.A. Brown H.S.	786	96
	Burke H.S.	1017	98
	St. John's H.S.	798	89
	Courtenay M.S.	525	99
	A.B. Rhett M.S.	476	97
	Rivers M.S.	545	92
	Haut Gap M.S.	450	97
Houston	G. I. Sanchez Alternative School	84	84
Virgin Islands	Elena Christian J. H.	1393	56
Lac Courte Oreilles	LCO Alternative School	92	70[c]
Plymouth	East Middle	860	25
	Central Middle	924	25
	Canton High	2350	49
	Salem High	2387	32
St. Paul	Murray J.H.	532	32
	Washington J.H.	702	0
	Johnson J.H.	1407	0
	Como S.H.	1094	0
	Central S. H.	1101	0

[1] No direct program services began during the first program year in the Miami, New Jersey and Milwaukee projects.

[2] This figure comes from the principal questionnaire, unless otherwise indicated.

[3] This figure is the number of students who were reported to be receiving direct program services at the time of the annual students questionnaire, unless otherwise indicated. Some projects such as Kalamazoo, St. Paul, and Charleston have school change as a major focus, and serve students indirectly.

[a] This is the number of students ever served. The school has a transient population.

[b] These enrollment figures are taken from school rosters or reports from the districts' Evaluation and Research Office.

[c] In addition, the project served 33 students at the Youth Center who were not enrolled in the alternative school.

[d] Only program participants surveyed. Total enrollment unknown.

Table 7

Ethnic Self-Identification of Students Receiving Services and the General Studentry in Project Schools

Project and group	Native Am.	Asian Am.	Span. Am.	Black	White	Other	N
Compton							
Receiving svcs.	0.0	0.0	19.4	80.6	0.0	0.0	36
Gen'l studentry	--	--	--	--	--	--	--
Pasadena							
Receiving svcs.	0.5	3.7	15.9	42.9	31.7	5.3	189
Gen'l studentry	0.5	4.0	17.9	42.8	30.3	4.5	201
Peer Culture Development							
Receiving svcs.	1.7	1.0	36.8	36.5	21.3	2.7	296
Gen'l studentry	0.6	2.2	44.0	23.2	28.3	1.8	505
Chicago Bd. of Educ.							
Receiving svcs.	3.3	0.7	35.5	43.4	13.2	3.9	152
Gen'l studentry	1.9	2.5	36.7	44.6	12.6	1.7	1113
Kalamazoo							
Receiving svcs.	--	--	--	--	--	--	--
Gen'l studentry	0.9	1.4	3.5	28.7	62.3	3.2	432
Bronx							
Receiving svcs.	1.1	2.2	35.5	61.3	0.0	0.0	93
Gen'l studentry	1.3	0.8	37.3	57.7	1.1	1.9	1398
Jazzmobile (Harlem)							
Receiving svcs.	1.3	0.7	5.3	91.3	0.0	1.3	150
Gen'l studentry	1.8	0.0	4.5	91.1	0.9	1.8	112
Puerto Rico							
Receiving svcs.	2.4	0.0	87.4	1.6	8.7	0.0	127
Gen'l studentry	1.2	0.2	91.4	1.6	5.5	0.0	813
Charleston							
Receiving svcs.	2.7	3.1	1.4	86.1	6.1	0.6	490
Gen'l studentry	1.3	0.1	0.4	83.6	13.9	0.7	1751
Houston							
Receiving svcs.	0.0	1.5	96.9	0.0	1.5	0.0	65
Gen'l studentry	--	--	--	--	--	--	--
Virgin Islands							
Receiving svcs.	2.2	0.0	22.2	71.1	0.0	4.4	45
Gen'l studentry	2.1	1.1	19.6	68.6	0.0	8.5	189
Lac Courte Orreilles	98.3	0.0	0.0	0.0	1.7	0.0	60
Plymouth							
Receiving svcs.	0.0	0.0	2.5	1.3	86.1	10.1	79
Gen'l studentry	3.3	0.0	0.5	2.2	89.0	4.9	182
New Jersey							
Gen'l studentry	1.2	0.6	19.4	19.1	54.1	5.6	1109
St. Paul							
Receiving svcs.	--	--	--	--	--	--	--
Gen'l studentry	2.0	0.8	1.7	10.5	84.5	0.6	894

Note.--Row percentages unweighted. These tallies are based on the student questionnaire file, thus excluding item non-respondents and students not taking the questionnaire, which was administered late in the Spring semester, 1981.

Table 8

Student-Reported Parental Education Levels for Students Receiving Direct Services and for the General Studentry in the Alternative Education Project

	General studentry[a]			Service recipients		
Project and school	Mean	SD	N	Mean	SD	N
Compton	--	--	--	4.39	1.95	28
Pasadena	6.21	2.07	190	6.10	1.94	171
Muir	5.92	2.21	100	6.31	2.18	80
Elliot	6.48	1.92	77	5.91	1.70	91
Peer Culture Development	*			*		
Chicago Board of Education	*			*		
Kalamazoo	5.21	2.23	386	--	--	--
Milwood	5.01	2.29	196	--	--	--
South	5.43	2.15	189	--	--	--
Bronx	4.58	2.45	960	5.12	2.43	59
22	4.10	2.56	102	3.25	3.20	4
55	4.94	2.62	66	6.50	1.73	12
63	5.95	2.38	43	**		2
64	4.69	3.07	78	2.88	2.59	8
82	4.42	2.45	93	5.43	2.15	7
117	3.98	2.32	111	4.80	2.28	5
132	4.78	2.60	41	5.11	2.42	9
145	4.64	2.27	88	**		1
147	4.89	2.33	45	7.25	0.96	4
148	4.64	2.18	39	**		2
166	4.76	2.46	107	4.80	2.12	5
229	5.53	2.19	116	--		0
Jazzmobile (Harlem)	*			*		
Otro Camino (Puerto Rico)	2.51	2.21	650	2.34	2.26	113
Santiago Gonzales	2.41	2.23	276	2.57	2.27	58
Ruiz Belvis	3.31	2.12	114	3.77	2.86	13
Dr. Alfredo Aguayo	2.26	2.15	252	1.60	1.77	42
Charleston	4.47	2.21	1416	4.48	2.22	385
242	5.06	2.14	249	--	--	--
741	5.01	2.12	167	4.95	2.37	64
742	5.44	2.23	121	5.58	2.24	62
743	4.69	2.17	121	4.84	2.34	50
751	4.77	2.06	191	--	--	--
754	3.64	1.87	147	4.19	1.98	53
755	4.18	2.01	135	4.41	1.85	63
944	3.73	2.30	161	3.62	2.03	66
951	3.22	1.94	124	3.00	1.71	27

Continued . . .

Continued

Project and school	General studentry[a] Mean	SD	N	Service recipients Mean	SD	N
Houston	--	--	--	1.92	1.77	54
Virgin Islands	3.31	2.29	115	3.38	2.16	29
Lac Courte Oreilles School	4.54	1.89	52	--	--	--
Lac Courte Oreilles Community Center	--	--	--	**		2
Miami	***					
Plymouth	5.59	1.69	179	4.89	2.07	72
East	5.62	1.59	52	4.44	2.22	16
Central	5.88	1.52	49	6.07	2.52	14
Canton	5.29	2.02	35	4.83	1.83	29
Salem	5.66	1.48	41	4.31	1.49	13
New Jersey	*			*		
Milwaukee	--	--	--	--	--	--
St Paul	*			*		

Note.--General studentry excludes direct service recipients. Students for whom project was known but school not known are shown in project total. Based on runs on the student questionnaire file, so excludes persons not responding to these questionnaire items or who did not complete the questionnaire.

[a]Weighted mean. Unweighted N.

*The questions upon which socioeconomic descriptions can be based were excluded from questionnaires for this project.

**Too few responses to report an average.

***No program participants.

--Not surveyed, or no persons in this category.

Table 9

School Action Effectiveness Study School Profiles:
Community and School Context

Project:	Community Crime	Gangs in School	Disruption	Victimization	Negative Peer Influence
Chicago Board of Education					
a		+	-	-	
b		+		-	
c			+		
d	+	+	+	+	
e	+				
f	+			+	+
g	+		+		
h	+	+			
i	+	+		+	
Bronx					
a					
b	+		+	+	
c		-		+	
d					
e	+		+		
f					
g	+	-	+		
n					
i					
j	-				
k					
l		-	+		
Jazzmobile					
a					
Puerto Rico					
a			-		-
b			-		
c			-	-	-
Charleston					
a	-			+	
b			+	+	
c			+		
d				+	
e	-			-	
f	-				
g	-	+		+	
h	-				
i	-	-	-		

Table 9 (cont.)

Project:	Community Crime	Gangs in School	Disruption	Victimization	Negative Peer Influence
Houston					
a		-	-	-	+
Plymouth<a>					
a	*	*	*		+
b	*	*	*		
c	*	*	*		+
d	*	*	*	-	
New Jersey					
a	-				-
b					
c			+	+	+
d			+		
St. Paul					
a	*	*	*	-	
b	*	*	*	-	
c	*	*	*	-	
d	*	*	*		
e	*	*	*		+
Compton					
a	*	*	*		+
Const. Rights Foundation					
a		+	*	+	-
b	-		*	-	*
Peer Culture Development					
a	+	+	-		
b		+	-		
c		+	-	-	
Kalamazoo					
a	-				
b	-				
Virgin Islands					
a	+	*			

<a>Sample sizes too small for dependable estimates.

Table 9 (cont.)

Project:	Community Crime	Gangs in School	Disruption	Victimization	Negative Peer Influence
Lac Courte Oreilles					
a	-	-	-		+

Note. Notation is as follows:
- A '+' indicates that the school average on this scale is above the 75th percentile for all schools in the study.
- A '-' indicates that the average is below the 25th percentile.
- A '*' indicates that the characteristic was not measured at the school.

Table 10

School Action Effectiveness Study School Profiles:
School Governance

Project:	St-tch Interaction	Planning	Fairness	Clarity	School Rewards
Chicago Board of Education					
a	-		-		-
b		-			
c	-	+			
d		-	-	-	+
e	-				
f	-		-		+
g	-		-	-	
h	-				
i	-			-	
Bronx					
a			-		
b		-			+
c	+		+	+	+
d		+			+
e					+
f		-		-	
g			+	+	+
h	+				+
i		+	+	+	
j			-	-	+
k			+	+	
l	+				+
Jazzmobile					
a					
Puerto Rico					
a	+	+	+	+	
b	+	+	+		+
c	+	+	+	+	+
Charleston					
a					
b		+	+		
c		+	+		
d					-
e		-			
f					
g		+			
h	-				
i					

Table 10 (cont.)

Project:	St-tch Interaction	Planning	Fairness	Clarity	School Rewards
Houston					
a	+	+	+	+	
Plymouth<a>					
a	-				-
b			-		
c		-		-	-
d	+			-	-
New Jersey					
a		-			
b	+				-
c	-	-	-	-	
d					-
St. Paul					
a	*	*		-	*
b	*	*	+		*
c	*	*			*
d	*	*			*
e	*	*	+		*
Compton					
a	*	*	+		*
Const. Rights Foundation					
a		-			
b	+	+	+		
Peer Culture Development					
a				-	
b				-	-
c	-			-	-
Kalamazoo					
a				+	
b				+	
Virgin Islands					
a	*	*			+

<a>Sample sizes too small for dependable estimates.

Table 10 (cont.)

Project:	St-tch Interaction	Planning	Fairness	Clarity	School Rewards
Lac Courte Oreilles					
a	+	+	+	+	

Note. Notation is as follows:
A '+' indicates that the school average on this scale is above the 75th percentile for all schools in the study.
A '-' indicates that the average is below the 25th percentile.
A '*' indicates that the characteristic was not measured at the school.

Table 11

School Action Effectiveness Study School Profiles: Student Composition

Project:	Alienation	Attachment to School	Belief	Involvement	Delinquency
Chicago Board of Education					
a	+		+		*
b	+				*
c				-	*
d		-		+	*
e			-	+	*
f	+	-	-	+	*
g	+	-	-		*
h				+	*
i					*
Bronx					
a			-		
b	+		-	+	
c			-		-
d	+	-	-	-	-
e	+	-	-		
f	+			-	
g					
h		+			
i		+	-		
j					
k		+			-
l					
Jazzmobile					
a			-		
Puerto Rico					
a	-		+		-
b	-	+	+		-
c	-	+	+	-	-
Charleston					
a				+	*
b		+		+	*
c		+			*
d				+	*
e	-	+	+		
f		+	+		+
g		+			
h	+				*
i					

Table 11 (cont.)

Project:	Alien-ation	Attachment to School	Belief	Involve-ment	Delin-quency
Houston					
a	-	+		-	+
Plymouth<a>					
a		-			*
b		-			*
c		-	+	-	*
d	-		+		*
New Jersey					
a		-			
b		-			+
c	+	-	-	-	+
d				-	
St. Paul					
a	-	+		*	*
b			+	*	*
c	-			*	*
d				*	*
e				*	*
Compton					
a	-	+		*	+
Const. Rights Foundation					
a	+	-			*
b	-	+	+	+	*
Peer Culture Development					
a	+				*
b		-	*	-	*
c				-	*
Kalamazoo					
a		-			
b		-			
Virgin Islands					
a		-	-	*	-

<a>Sample sizes too small for dependable estimates.

Table 11 (cont.)

Project:	Alien-ation	Attachment to School	Belief	Involve-ment	Delin-quency
Lac Courte Oreilles					
a					+

Note. Notation is as follows:
- A '+' indicates that the school average on this scale is above the 75th percentile for all schools in the study.
- A '-' indicates that the average is below the 25th percentile.
- A '*' indicates that the characteristic was not measured at the school.

Table 12

Indicators of Commitment to Institutionalization of System Change, and of Potential for Public School System Adoption

1. Institutionalization or system change is a stated goal of the project, and the project has a strategic plan for institutionalization or system change.

2. Project sees itself and is regarded by others in its environment as a demonstration or pilot project.

3. Project maintains contact with public school system.

4. Key public school system decision makers have influence in project planning, implementing, and evaluation and review.

5. Project staff influence public school decision making in planning, implementing, and evaluation.

6. The project provides systematic feedback to public school system on project progress and activities.

7. Project is developing methods or models that can fit within the public school structure.

8. Project costs per student are the same or lower than the system's current costs per student.

9. System has a stated policy or goals regarding alternative education and has made a commitment to integrate project or key elements.

10. System has experience with adopting and integrating innovations; one or more key decision makers has a reputation for innovation and experimentation.

11. System is in crisis and seeking answers.

12. System has assigned someone responsibility for monitoring project or receiving information from it, and has established a vehicle for dissemination.

13. System provides budget, personnel, and services (transportation, lunch, testing, etc.).

14. System sees project personnel as "insiders" rather than inexperienced or invading "outsiders."

Subject Index

Name Index

Name Index

Part II

Project Narratives

Abstracts of Part II Chapters

The following pages contain brief abstracts of the chapters in Part II of the School Action Effectiveness Study's first interim report. Single copies of these chapters are available from the Center for Social Organization of Schools while the supply lasts. When requesting a chapter, please specify the chapter's title and author.

Interim Evaluation of Project PATHE--Charleston

Denise C. Gottfredson

Abstract

PATHE--Positive Action through Holistic Education--is a Charleston County School District in-school project aimed at reducing delinquency, increasing attendance, increasing postsecondary school attainment, and increasing academic achievement. The underlying philosophy of the project is that an integrated approach is necessary to effect changes in student behavior and attitudes. The project organizes administrators, faculty, staff, students, parents, and community leaders in planning and implementing strategies to solve the problems of the Charleston schools. PATHE gives individualized affective, academic, and vocational services to all youths in the PATHE schools, although it focuses on a group of 100 students per school especially in need of the project services. In addition to direct student services, PATHE provides training and resources to teachers and works toward organizational-level changes in policy and procedures. Absence of an evaluation design for the 1980-81 school year makes a rigorous assessment of the project's effectiveness impossible. This report analyzes implementation data and identifies the project's weak and strong components.

Academy for Community Education: Interim Report

Deborah Daniels

Abstract

The Academy for Community Education (ACE), Miami, Florida, is the alternative education project run by the Institute for Innovative Interventions, Inc. The Institute is a not-for-profit corporation formed in 1980 by a group of concerned Dade County citizens. Although the Academy is an independent alternative school, it can share Dade County Public School system resources such as teaching personnel. Its participants are selected primarily from the disadvantaged area of Coconut Grove, although it is open to other students as well. The school is mainly designed to serve pre-delinquent youths who have demonstrated truancy, disruptive classroom behavior, excessive tardiness, and low levels of academic achievement. The Academy's main purpose is the development of successful approaches for working with these students that may be incorporated into the Dade County public alternative schools, as well as the alternative programs of other school systems. The project was funded in January, 1981, and began services to participants the following summer. This narrative describes the project's planned interventions and its start-up activities.

Peer Culture Development (PCD), Chicago

Jane St. John

Abstract

Peer Culture Development (PCD) is an intervention in several Chicago public schools aimed at decreasing delinquency, improving attendance, increasing achievement, and altering school disciplinary practices. PCD assumes that peer culture in some instances generates a set of subcultural values that are counterproductive in a school environment, and that schools have not always been able to help students subscribing to these values. The project therefore attempts to harness peer pressure to alter student values and behavior, and to implement school procedures that will redirect students. Some students meet daily in small classroom groups as part of their regular school program to help each other solve problems, with the guidance of a PCD counselor and a set of straightforward and clearly articulated values. Other students in the school are referred to these classes for crisis intervention. Preliminary evidence indicates that the project is being well implemented, has plausibility, and shows early evidence of effectiveness in some areas. More evidence will be required for strong claims of effectiveness because large sample sizes are not yet available. The project involves a rigorous evaluation component and is continuing to implement that component, so reasonably conclusive evidence should be forthcoming during the next year.

Otro Camino, La Playa de Ponce, Puerto Rico: Interim Report

Jane St. John

Abstract

Otro Camino was established to provide a supportive environment for the youth of three schools in La Playa. The project staff members hoped to demonstrate that student interest in learning could be captured by providing tailor-made activities for each student. They predicted that their interventions would reduce student alienation and result in higher levels of academic and vocational attainment. They also expected that their project would encourage students to stay in school and would discourage vandalism. Initial assessments imply that the project was reasonably well implemented, but the staff has decided to change major components because explorations uncovered the approaches that students especially liked. The project's evaluation design will allow an assessment of effectiveness by the end of its second year of operation.

Project PREP: An Interim Report of its Evaluation

Deborah K. Ogawa

Abstract

Project PREP serves 12 schools in School District 9 in the South Bronx area of New York City. By meeting students' academic and emotional needs, and by increasing parent, teacher, and student involvement in school activities, Project PREP anticipates attenuating disruptive behaviors in the schools. The interventions include an Alternative School where four feeder junior high schools refer students; and eight Citizenship Cluster Schools, where 15 students from each school are selected to participate in a traditional school environment with non-PREP students. In addition, all Project PREP students can participate in after-school, evening, and Saturday activities through a Youth Program. The comparison of PREP students with a random sample of 300 students from each school showed no significant difference between the two groups. Steps required to strengthen the evaluation of this project in its second year are described.

The Plymouth-Canton Alternative Education Project: Interim Report

Richard Carlton

Abstract

The Plymouth-Canton project serves two high schools sharing a common large campus and two middle schools in an area on the outer rim of the Detroit Metropolitan area. It extends earlier in-school counseling services and an out-of-school program for students experiencing difficulty in the large high school environment, by providing similar services in the middle schools and by developing a high school remedial writing and study skills component. The project began its first semester of operation in spring 1981 without provision for evaluation, and is now attempting to implement its project in a partially evaluatable form during the 1981-82 school year.

Student Training Alternatives through Urban Strategies (Project STATUS): Interim Report

Richard Carlton

Student Training Alternatives through Urban Strategies (Project STATUS) is an effort to combine and further develop two existing experimental program models: one, a citizenship (social studies and English) curriculum which draws heavily on law-related education materials designed by the Constitutional Rights Foundation, and the other a youth involvement and school climate improvement strategy. Both models have received considerable developmental and implementation work over the past decade. Project STATUS believes that young people tend to rebel against both school and society and to get in trouble for any combination of the following reasons: (a) The young people do not understand or believe in the legitimacy of the legal and authoritarian structures of society's institutions, (b) They are not involved in relevant, meaningful educational programs. Or, (c) they are excluded from participating with adults in the decision-making structures of the school. The first year of operation of this project involved developing the interventions and staff skills, and implementation was impeded to some extent by the timing of project funding--after school had begun in the fall. It now appears that a more complete implementation of Project STATUS is likely in the upcoming year, and that prospects are good that at least some elements of the project will be evaluatable.

The George I. Sanchez School: Interim Report

Deborah Daniels

Abstract

The George I. Sanchez Alternative Education Program (GIS), is the alternative education component of the Association for the Advancement of Mexican-Americans (AAMA). Introduced in 1973 as a half-day alternative education program, the project today is an accredited junior and senior high school, emphasizing services to Hispanic youth, ages 12-18, who have dropped out or appear to be on the verge of dropping out of public school. Its students are drawn from referrals from seven Houston Independent School District (HISD) schools, participants or former participants in other AAMA programs, and referrals from other social service programs. The school seeks to provide an educational environment in which young people become full partners with school staff in determining objectives and strategies for achieving them. Individual educational planning and student and parental participation in school decision-making and implementation are emphasized. The project also conducts a program of after-school recreational activities and educational and occupational counseling. This after-school component is referred to as the Alternative Activities Program (AAP). No statements about project effectiveness can be made at present, but the project can be described. Plans to make some aspects of this project evaluatable in the upcoming year have been developed.

The Milwaukee Youth Employment Center

Carol Yamasaki

Abstract

The Milwaukee alternative education project, the Milwaukee Youth Employment Center (MYEC), is an education and employment program implemented by the Jewish Vocational Services of Milwaukee, Inc. (JVS). The project attempts to bring together the resources of existing youth-serving agencies within the city in a shared effort to address the needs of youths 16 and 17 years old who have dropped out of school. A consortium of agencies refers youths to MYEC for individualized instruction and counseling aimed at preparing them for employment.

The project has an additional school component. This includes a Return Center within the public school system, designed to assess the needs of dropouts and place them in appropriate educational options. It also includes the integration of the competency-based, work-related curriculum designed and implemented at MYEC into the existing work-study curriculum at six public schools.

The project has adopted the program development process as an integral part of its operation. All staff members participate in the use of the model for project planning and for tracking progress. The project has implemented a true experimental design: Staff randomly assigns youths to the program and will have at the end of the year a control group of 150 youths for comparison.

The Compton Action Center for Youth Development Alternative School: Interim Report

Deborah Daniels

Abstract

The Compton Action Center for Youth Development (CAYCD) Alternative School is the sole alternative program serving junior and senior high students from the Compton Unified School District (CUSD). Originally funded as an individual and family treatment program for identified assaultive youth referred by the criminal justice system, the project evolved first into a mini-school and then into a year-long alternative school. Today its target population and referral base have broadened, although most students are referred from CUSD. Participants range from youths who are heavily involved in gang activity or who have been arrested a number of times to youths with little previous history of trouble in or out of school. All its clients share a sense of alienation from, frustration with, and poor adaptation to traditional learning environments.

The CACYD Alternative School seeks to establish an environment in which students and parents feel that they are an integral part of the planning for and implementation of the student's education. Students in the alternative school are carried on the CUSD attendance rolls and receive academic credit through CUSD. The alternative school program meets all of the CUSD requirements for graduation. CACYD, however, seeks to create a flexible learning environment, one which will accommodate individual student interests and needs. The project aims to do this through (a) computer-assisted instruction, and (b) individual education plans, which establish academic and behavioral objectives and strategies to achieve these objectives.

The effectiveness of this project during the past school year (1980-81) cannot be determined. Project personnel are taking steps to make the project more evaluatable next year.

Project RETAIN, Chicago Board of Education: Interim Repost

Jane St. John

Abstract

RETAIN (Responsive Education Through Alternative Instructional Networks) is aimed at the problems of poor attendance, disruptive behavior, and low achievement in Chicago public schools. Attendance problems include class-skipping and absenteeism in the high schools and absenteeism in the elementary schools. Disruptive behavior includes minor classroom or school disruption as well as serious assaults. Achievement problems are evident in the results of standardized tests. The project considers poor attendance and low achievement both individual- and school-level problems. The primary focus of the project is the implementation of Individual Learning Plans (ILP's). The project's staff reported that ILP's developed in the early stages of implementation were not used effectively. Accordingly, they decided that more in-service training for RETAIN techers was needed to assist them with developing and using this approach to education. Plans for the in-service training were carried out in September, 1981, just as school opened. The project is making efforts to strengthen its intervention in its second year, and there is every reason to believe that a stronger program will have a stronger impact on the students it serves.

The Milwood Alternative Education Project

Richard Carlton and Michael Cook

Abstract

The Milwood Alternative Education Project is a collaborative effort between Western Michigan University and the Kalamazoo Public Schools. The project attempts to reduce delinquent behavior and improve attendance and achievement through modifications in the policies and practices of a single school. Interventions include monitoring and follow-up on attendance problems, an in-school suspension room, staff development, tutoring, student and community involvement strategies, and task forces to design changes in several areas of school operation.

The project attempts to prevent delinquency by changing the school at the structural and programmatic level, and by altering school policies and procedures. In particular, it seeks to make changes in the school reward structure, response to non-attendance, and response to discipline problems. It tries to increase the participation of students in

school activities--especially activities designed to increase students' feelings of involvement in and attachment to school.

The Lac Courte Oreilles Alternative Education Project: Interim Report

Richard Carlton

Abstract

This project focuses on two groups within the Indian youth population on the Lac Courte Oreilles (LCO) reservation. The primary target is youngsters who have dropped out of a nearby public school system, but who have not enrolled in the LCO system. The secondary target group is students enrolled in the LCO High School who do not attend regularly enough to keep up with the school program and are suspended for nonattendance. The project seeks to enroll these two groups in its alternative education project, and to provide opportunities for reservation youth to engage in productive activity evenings and weekends by providing educational activities during those periods. Several difficulties which surfaced over the past year will have to be successfully resolved if the project is to continue to develop. First, fuller implementation will require that the Youth Centers extend their hours of operation to the evenings and weekends. Second, more vigorous outreach efforts will apparently be required if the project is to enroll a substantial number of youths who are not enrolled in other educational programs. Third, improved methods for keeping track of the users of the AEP services, and a method to learn what happens to them will be needed to assess the AEP activities and to further develop them.

The Virgin Islands Alternative Education Project

Jane St. John

Abstract

This project aims to increase teacher competencies in alternative education techniques and student self-esteem to increase student academic success, reduce class skipping, decrease alienation and decrease delinquency. Major categories of interventions planned by the project to reach these goals and objectives include: (a) training in alternative education techniques for teachers, (b) providing intensive exposure to alternative education classrooms for 60 students, (c) holding a public relations campaign for the project, and (d) arranging activities that allow the participa-

tion of parents and community members in the program. The program is intended to break the cycles of strain and disorganization by training teachers in methods that should help youths experience success and stay on the "straight and narrow" path.

New Jersey Educational Improvement Center-South

Donald E. Rickert, Jr.

Abstract

The key to the EIC-South's theory of deliquency is youths' involvement and participation in the things affecting them. The overall thrust of the EIC-South Alternative Education Project is to influence school climate and programs, and to establish a community problem-solving process for reducing delinquency and its associated problems. Five primary components will be established: (a) a school climate improvement component, (b) a community process component, (c) a youth participation component, (d) a public relations component, and (e) a leadership and training resources (technical assistance) component. This project was funded several months later than other projects in the Alternative Education Program. This made the period between January and September, 1981, essentially an extended planning and start-up period.

The Jazzmobile Alternative Arts Project: Interim Description

Donald E. Rickert, Jr.

Abstract

The Jazzmobile Alternative Education Arts Project aims to utilize the arts as a medium through which juveniles can constructively channel their energies. The project is intended to deliver an arts-oriented program to juveniles in sixth, seventh, and eighth grades who show disruptive behavior, who are chronic absentees and truants, or who experience academic failure. Perhaps partly because a rigorous evaluation was not anticipated by Jazzmobile or the school system, the evaluation of this project has encountered repeated difficulties.

Project Together: Interim Report

Gary D. Gottfredson

Abstract

This project is designed to address three related systemic problems faced to some degree by nearly all schools: (a) a failure to develop the practical, real life skills students need, (b) student dissatisfaction, boredom, and non-attachment to school, and (c) low public regard for the schools. The theory underlying this project assumes that system changes will depend upon making broad changes in many of the structural arrangements and school practices that contribute to the problems identified earlier. This theory implies that multiple interventions aimed at organizational change will be necessary to bring about changes in (a) the management of learning by teachers in the classroom, (b) student competencies in managing interpersonal relations, (c) behavioral control methods used in the classroom, (d) the breadth and extensiveness of use of community resources, (e) methods used to establish and enforce rules in the school, and (f) the extensiveness of parent involvement in school decision-making or interaction with school personnel. The interventions being implemented include: (a) Action Learning Projects undertaken on student initiative to enable them to solve real world problems and in which teachers serve as managers of learning rather than as instructors, (b) staff training, (c) peer counseling, (d) internships, apprenticeships, volunteer activities, and field trips, (e) FOCUS, (f) media efforts, (g) administrator meetings and in-service training, (h) parent-teacher training on adolescent needs, and (i) advisory councils of parents, students, and staff.

END