*ISSUES IN MULTIVARIATE ASSESSMENT OF A
LARGE-SCALE BEHAVIORAL PROGRAM*

JAMES FILIPCZAK, MARGARET B. ARCHER, MICHAEL S. NEALE,
AND RICHARD A. WINETT

INSTITUTE FOR BEHAVIORAL RESEARCH, INC.

Several social and research issues directly affected the development and implementation of multivariate assessment in a large community-based applied research program. Examples are drawn from experiences of the Preparation through Responsive Educational Programs Project for disruptive and skill deficient adolescents in suburban, rural, and urban junior high school settings, focusing on the assessment of academic and social skill development and long-term skill maintenance. The social context altered both project treatment and follow-up plans, requiring assessment of potentially unintended effects and decreasing consistency across sites. Future community acceptance of such programs may depend on the investigators' adaptation to diverse community pressures for program conduct and assessment and the measurement of phenomena that are not always directly observable.

DESCRIPTORS: academic behavior, social behavior, evaluation, school-based training, community setting, adolescents

Applied behavior analysis has been typified by research designs with few or single subjects, highly controlled treatment conditions, few or single response measures, and visual analysis of outcomes through graphing (see issues of *Journal of Applied Behavior Analysis*). However, the use of multiple behavioral treatment and assessment procedures is increasing where behavior enhancement of several clients or groups is attempted simultaneously—for example, the

prison (Cohen and Filipczak, 1971; McKee and Clements, 1971) and the school (Filipczak, Friedman, and Reese, 1979; Heaton, Safer, Allen, Spinnato, and Prumo, 1976). In these programs, single-subject strategies may not be entirely appropriate or responsive to larger social concerns (Yates, 1976).

This article reports on ways in which the Preparation through Responsive Educational Programs (PREP) Project developed and conducted process and outcome assessment in a behavioral program that incorporated several treatment variables. The intent is to show, through pertinent examples, the continuity that was maintained and the practical constraints that were involved as PREP attempted to respond to changing student learning needs, increased research demands, and emerging evaluation responsibilities.

During the seven-year research and demonstration history of PREP, its major concern has been with outcomes rather than with the isolation of efficacious variables (Dietz, 1978). However, PREP has regularly undertaken internal process and outcome assessment, and it has been subjected to systematic evaluation from outside

The research reported here was funded by Grants MH-14443 and MH-21950 to the Institute for Behavioral Research from the Center for Studies of Crime and Delinquency of the National Institute of Mental Health, and Grant 5S07-RR05636 from the Division of Research Resources of the National Institutes of Health. Other substantial support was provided by local cooperating school districts and personnel. The authors would like to thank all former PREP students, parents, teachers, and staff (anonymously) for their effort that made this report necessary. Special debt is also acknowledged to Harold L. Cohen and Israel Goldiamond, who saw the appropriateness of helping children learn to help themselves. The views expressed are those of the authors and do not necessarily represent views or policies of any agency named here. Requests for reprints should be sent to James Filipczak, Institute for Behavioral Research, 2429 Linden Lane, Silver Spring, Maryland 20910.

sources. The types of issues and problems generated by multiple audience criteria are addressed here.

Finally, many of PREP's procedures were formulated during the infancy of behavioral technology and have since been refined or modified in accordance with the "start of the art." Therefore, assessment in PREP must be viewed as an emerging program that has evolved over time.

PREP GOALS AND CONSTRAINTS

PREP hypothesized that the combined effect of short- and long-term changes it supported would help establish behaviors incompatible with major juvenile problems. In the short range, PREP sought to improve students' academic and school-related social skills, proposing that increments in students' rate and accuracy of school work would yield increased access to reinforcement available in the school environment and increased contact with socially appropriate models. Reciprocally, PREP tried to enhance the school's ability to provide positive incentives for students' participation and progress. Over the longer range, PREP posits that student academic and social skill improvement, increased access to acceptable social roles, and reinforcement by pro-social persons will lead to expanded incentive for further socially acceptable behavior.

PREP's long- and short-range assessment efforts were governed by three forces: (a) the breadth of these conceptual goals, (b) the limitations imposed by often conflicting concerns of different audiences, and (c) the constraints within the settings and measurement instruments themselves.

The choice of, and priorities given to, measures by PREP and its outside evaluators reflect the diverse interests that an applied research program must satisfy to be deemed successful or meaningful (Wolf, 1978). PREP audiences have included its funding agency, school administrators, juvenile justice officials, teachers,

parents, students, reviewers from the fields of psychology, education, criminal justice, and community services, governmental investigative agencies, and local news and special interest groups. Furthermore, PREP's adherence to its treatment goals constituted still another "audience" (program staff) by which to measure program success.

Like other programs in natural settings, PREP's selection of specific dependent measures and treatments occurred within a number of constraints—system priorities, program feedback, short-term outcomes, times and cost limitations, and legal and ethical issues. Administrators, teachers, students, and parents at each site presented their own preferences for, and obstacles to, the implementation of certain procedures. Feedback from ongoing or recently completed programs also suggested major adjustments in procedure and design. In addition, time and cost limits influenced the program's form between implementation and follow-up. Even where cooperation of students and parents was obtained, legal and ethical issues typical of the 1970's emerged as staff sought access to school and court records.

THE PREP PROGRAM

Treatment

The PREP program encompassed one school year. PREP dealt with voluntary consenting junior high school students who presented the most severe academic and/or social problems to their schools during the prior year. These students were recommended by school staff on evidence of failing grades, numerous disciplinary referrals, high absenteeism, low performance on standardized tests, or contacts with police.

Over the nine years reported here, PREP was conducted in a laboratory school and in a suburban, a rural, and an urban public junior high school. Between 30 and 120 students enrolled each year. Table 1 describes the setting and experimental designs for each of these years.

Table 1
Setting and Experimental Design for Each PREP Year

<i>Year</i>	<i>Setting</i>	<i>Experimental Design</i>		<i>N</i>
1969-70	Laboratory	1 Group:	● Academic Plus Social	15
1970-71	Laboratory	1 Group:	● Academic Plus Social	15
1971-72	Suburban	2 Groups:	● I—Academic Plus Social	30
			● II—No Treatment Control	30
1972-73	Suburban	5 Groups:	● I—Complete Program	16
			● II—Academic Only	16
			● III—Social Only	16
			● IV—Family Only	16
			● V—No-Treatment Control	16
1973-74	Rural	2 Groups:	● I—Academic Only	39
			● II—Social Plus Family	39
1974-75	Rural	3 Groups:	● I—Academic Only	39
			● II—Social Plus Family	36
			● III—No-Treatment Control	32
1975-76	Rural	Factorial:	● Factor I—Academic versus No Academic	48
			● Factor II—Social versus No Social	48
1974-75	Urban	2 Groups:	● I—Complete Program	47
			● II—No-Treatment Control	36
1975-76	Urban	2 Groups:	● I—Complete Program	60
			● II—No-Treatment Control	60

PREP's program (Filipczak, Breiling, and Storm, 1973; Institute for Behavioral Research, 1972*b*) included: academic training that permitted rapid skill development in reading, English, and mathematics; and social or interpersonal skill training that facilitated generalizable behavioral skill development. Family and teacher training comprised smaller, yet vital, components in all settings.

Assessment began when students enrolled, using selection information as the basis for gauging student performance changes over the program year. PREP also administered achievement and/or diagnostic tests for assessing academic change.

Each PREP student received intensive daily instruction in English, reading, and mathematics. Self-instruction was stressed with certain objectives organized for small-group work. After diagnosis and prescriptive placement, students worked in pertinent topics and skill levels of self-instructional programs on a continuum of 2nd through 11th grade objectives. The Indi-

vidually Prescribed Instruction (IPI—Appleton-Century-Crofts) was one such program that used continual assessment through curriculum embedded tests. Data were also maintained about the students' rate of work, on-task activities, and social behavior.

If specified by the experimental design, students participated in a social skills training class. Here, the teacher assisted the students in learning skills for dealing with school, home, and community through a curriculum promoting appropriate interactions with school staff, parents, peers, and persons of authority pertinent to each setting. Students engaged in problem solving for in-school and community situations and developed skills in nonverbal behavior, small-group interaction, self-control, instruction following, and test taking.

Contingency management procedures of varying complexity were used in both academic and social components. Students earned both social and tangible reinforcement for attaining specified academic and behavioral objectives. Regular

feedback on student progress to both students and parents was also an integral part of PREP.

Students in PREP attended four or five regular school classes per day, such as science or physical education. This helped ensure that students were less labeled by PREP participation and provided them with skills not a part of PREP training. This permitted staff to assess the extent to which the skills learned in PREP carried over to other settings, using non-PREP class teachers to rate student academic and social skills.

Follow-up

PREP's original planning did not include long-term follow-up. However, the importance of assessing student skill maintenance after treatment was soon recognized as a major ingredient of cost-benefit analysis by staff, the funding agency, and school personnel.

To date, information has been gathered up to four years after treatment for the laboratory, suburban, and urban school groups. Follow-up data collection has recently been completed at the urban site, and rural school follow-up awaits implementation during the upcoming year.

Selecting measures appropriate to PREP's long-term purposes and reconciling these with student time, ethical constraints, funding limitations, and public school hesitancy gave the project substantial conceptual and practical problems (Filipczak, McCombs, Rusilko, and Koustenis, Note 1).

Between 67% and 80% of the original groups participated in each completed follow-up. Most were high school juniors or seniors, and analyses showed the follow-up samples were representative of the original matched groups. Students were located by mail and telephone if they had left school or if access to school records was prohibited. Self-report measures and achievement tests were given in special sessions held at school or at community facilities. All subjects who completed the self-report were offered a nominal gratuity. Confirmations of

self-report were obtained by telephone from the student's mother or other closest relative.

The self-report questionnaire contained elements compiling the subject's personal status and demographic information, experience in and benefits from PREP, involvement in leisure activities, self-esteem, aspirations and expectations, involvement with juvenile justice agencies, relationships with family and friends, anticipated consequences of contacts with juvenile authorities, and personality adjustment.¹ Data gathered from school records were treated as raw scores for clear interpretation. Achievement tests used at follow-up repeated those administered during the program.

The analyses of process and outcome measurement in PREP are reported here through references to original data in both published and unpublished documents. Reports on all treatment years and sites are included. Once PREP was implemented in the public school, often more than 20 different process or outcome measures were used each year. Only a limited number are reported here, the sample intending to reflect the various rationales for selecting and terminating use of measures.

ACADEMIC MEASURES

Each year, PREP staff used several measures of student academic performance to assess teaching-learning processes and overall project outcomes. At the four-year follow-up, certain of these measures were repeated in conjunction with longer-range measures of student academic attainment

Program Measures—Process

Assignment completion. For all PREP classes, assignment completion was tabulated routinely. However, both the basic unit of measurement and the time period before feedback varied across sites (Table 2).

¹Copies of the actual instruments used in PREP follow-ups are available from the first author.

Table 2

Consistency and Change in Selected Academic Assessments by Site and Year

Assessment	Laboratory			Suburban			Rural			Urban		
	1969-1971	1971-1972	1972-1973	1973-1974	1974-1975	1975-1976	1974-1975	1975-1976	1974-1975	1975-1976	1974-1975	1975-1976
ASSIGNMENT COMPLETION:												
Data Source	Daily Records Within A	Daily Records Within B	Daily Records Within C	Weekly Records Within D	Weekly Records Within E	Weekly Records Within F	Weekly Records Within G	Weekly Records Within H	Quarterly Records Within G	Quarterly Records Within H	Quarterly Records Within	Quarterly Records Within
Citation ^a												
GENERAL ACADEMIC ACHIEVEMENT:												
Data Source	SAT Tests	SAT Tests	SAT Tests	Diagnostic Tests	SAT Tests	SAT Tests	SAT Tests	SAT Tests	SAT Tests	SAT Tests	SAT Tests	SAT Tests
Design ^a	Within*	E vs. C*	Mult*	Mult*	Mult*	Mult*	Mult*	Mult*	E vs. C*	E vs. C*	E vs. C*	E vs. C*
Citation	I	J	K	L	M	N			N	N	N	N
READING ACHIEVEMENT:												
Data Source	Gates Test Within*	Gates Test E vs. C*	Gates Test Mult									
Design ^a		J	O									
Citation	I											
CLASS GRADES:												
Data Source	All Classes Within I	All Classes E vs. C*	All Classes Mult*	All Classes Mult*	All Classes Mult*	All Classes Mult*	All Classes Mult*	All Classes Mult*	All Classes Mult*	All Classes Mult*	All Classes Mult*	All Classes Mult*
Design ^a		J	P	Q	Q	Q	Q	Q	Q	Q	Q	Q
Citation												

* Differences assessed and significance obtained in favor of treatment group(s).

^aDesign Code: Within = Within-Group Comparisons, E vs. C = Experimental and Control Comparisons, Mult = Multiple Group Comparisons

Citation Legend:

- A. Cohen *et al.*, Note 2.
 B. Cohen *et al.*, 1974.
 C. Friedman *et al.*, Note 3.
 D. Picke and Filipczak, Note 4.
 E. Fiordaliso *et al.*, Note 6.
 F. Fiordaliso and Archer, Note 5.
 G. Filipczak and Friedman, Note 7.
 H. Filipczak *et al.*, Note 8.
 I. Cohen *et al.*, 1971.
 J. Filipczak, Note 9.
 K. Friedman *et al.*, Note 10.
 L. Filipczak and Wodarski, Note 12.
 M. Filipczak *et al.*, Note 13.
 N. Filipczak *et al.*, 1979.
 O. Weinberg and Breiling, 1977.
 P. Filipczak *et al.*, Note 14.
 Q. Friedman *et al.*, 1977.
 R. Reese and Filipczak, *in press*.

In the laboratory setting, project staff evaluated all instructional materials to determine reading and other response requirements through an objective formula for "work units" in each completed self-instructional task. Time measures were also kept for each phase of study, yielding a work units/hour value used to assign student grades and token system payment. Calculations were computer based and printed records were available daily (Cohen, Filipczak, Slavin, and Green, Note 2).

PREP continued this computer-based work unit feedback in the suburban school, even though it seemed impractical for the natural setting. It proved costly, and, during the second project year, both school and project staff recommended against continuing the daily system, despite its positive effects on academic and behavioral performance (Cohen, Filipczak, Boren, Goding, Storm, Bishop, and Breiling, 1974; Friedman, Filipczak, Fiordaliso, and Picek, Note 3).

At the rural school, teachers and community reviewers concurred that the complex work unit system did not appear transportable from the laboratory (Picek and Filipczak, Note 4). The assignment completion measure was simplified to the number of pages completed. Though daily record charts were posted in classes, summaries were only computed weekly (Table 2) for review by teachers and staff. Class management did not appear to be harmed by this change. In fact, direct teacher participation improved its relevance, permitting comparative monitoring of different classes during weekly teacher seminars (Fiordaliso and Archer, Note 5), and measurement of special behavior program outcomes (Fiordaliso, Friedman, Brotman, and Lordeman, Note 6).

The urban school teachers preferred less complex compilation of assignment completions, counting as an assignment any unit of work that could be reasonably completed in one class period. Teachers posted a daily record of assignments completed, informing parents about the totals on a quarterly basis (Filipczak and Fried-

man, Note 7). Professional visitors commented on the utility and simplicity of this procedure, which permitted the efficient management of a multiclass incentive system (Filipczak, Reese, Fennell, Bass, Kent, and Gilmore, Note 8).

The lack of consistency in measuring assignment completion across sites did not appear to weaken the support provided the teachers and students. Each site selected or developed its own effective tactics for dealing with this functional measure of student academic performance.

Program Measures—Outcome

General academic achievement. In PREP for all but one program year general academic achievement was measured. The Stanford Achievement Test Battery (Intermediate Levels I and II, 1964 Edition, Forms W and X; or Intermediate Levels I and II, 1972 Edition, Forms A and B) served as a nationally accepted measure of student skills through its subtests in mathematics, language arts, and reading comprehension. Each year, selected subtests were administered twice or more to all participant groups, permitting treatment comparisons. PREP selected the Stanford Achievement Test Battery (SAT) during the laboratory years because its content approximated the project's curriculum objectives, and it did not conflict with the three cooperating school systems' yearly testing programs. Over the laboratory years SAT results showed PREP students making significant 1.5 to 2.0 grade level gains over their entry scores on most subtests (Cohen, Filipczak, Slavin, and Boren, 1971).

County administrative personnel for the suburban school sanctioned the SAT because it linked objectives used in both the PREP and the county curriculums. For the first suburban year, experimental students significantly outperformed controls on all but one subtest of the SAT (Filipczak, Note 9). The test was retained for use the second year with the endorsement of school and project staff. During a controversy over the use of behavior modification in the public schools, the school board president

demanding that PREP conduct additional mid-year academic testing. At posttesting, many students voiced or demonstrated dissatisfaction with the several testing sessions, which may have been reflected in the general decrement in test performance for both experimental and control groups that year (Friedman, Archer, and Lorde-man, Note 10). Both academic skill assessment and assignment completion data noted no substantial differences between the first and second year. Later academic assessment at follow-up indicated experimental group superiority, confirming the process measures (Filipczak, Neale, Winett, and Rusilko, Note 11).

At the rural site, teachers and administrators substituted diagnostic tests for achievement measures (Filipczak and Wodarski, Note 12). The absence of norm referencing, however, precluded use of this diagnostic test for achievement assessment functions. PREP teachers and staff agreed to use the SAT subtests during later years at the rural site. In each of these years, the SAT provided confirmation of observed process changes and demonstrated academic treatment group superiority over both social and control groups (Filipczak *et al.*, 1979; Filipczak, Friedman, Fiordaliso, and Picek, Note 13).

Because the urban students showed very low academic skills at entry (Filipczak and Wodarski, Note 12), PREP staff and teachers concluded that only the reading and mathematics computation subtests of the SAT would be used. These subtests served as brief overall achievement measures and references to an appropriate placement test level in language, reading, and mathematics. Again, their use as comparative measures of achievement was justified in each year's outcome analysis, but the students' lack of adherence to testing rules at the end of each year provided an added source of error to the test situation (Filipczak *et al.*, 1979).

References noted in Table 2 reflect that the use of the SAT over seven different site/years was repeatedly justified by its ability to discriminate successfully among treatment groups and to provide accurate general placement for

students in various levels of specific program placement testing.

Reading achievement. In the laboratory and suburban settings, the Gates MacGinitie Reading Test (1965 Edition, Survey E, Forms 1M and 3M) was used to assess reading achievement. This instrument included measures of reading speed, accuracy, and vocabulary in addition to a test of comprehension like that provided in the SAT. During these early years, the PREP language curriculum did not include sophisticated placement tests like those available through the IPI mathematics sequence. Consequently, PREP and suburban school staff concurred that the assessment of change in academic performance and the accurate instructional placement of students demanded a multi-purpose reading test like the Gates.

In both settings, the Gates provided outcome testing relevant to the objectives and content of the PREP instructional program. It also assessed within-group, pre-to-post change (Cohen *et al.*, 1971) in the laboratory years and revealed differences between groups, significantly favoring experimentals in the first suburban year (Filipczak, Note 9). Again, the disrupted testing conditions during the second suburban year may have militated against accurate performance assessment for all groups (Weinberg and Breiling, 1977).

After process measures comparable to the IPI in mathematics were found for reading, the Gates was not recommended by PREP staff for either the rural or urban sites. School teachers and administrators concurred, returning to the unidimensional SAT comprehension subtest. The Gates was used later, however, in the follow-up for the suburban school.

Class grades. Originally, class grades were thought to be questionable measures of student academic skill in school because of the subjectivity of their underlying units and the extent to which they incorporate social behavior. However, during the laboratory and suburban years, both parents and teachers convinced PREP staff that positive labeling of students could be

achieved if their class grades improved, and that the interaction of academic scores and behavioral performance was indeed a practical reflection of academic performance. It was also clear that grades were low-cost, unobtrusive, and nonreactive measures, and that the bias in such a measure would likely be distributed equally across the randomly assigned groups.

At the laboratory site, class grades served as a within-group measure of the student's participation in regular school classes; however, no systematic analyses were conducted with these data (Cohen *et al.*, 1971).

Measuring internal program progress and the generalization of student skills to non-PREP class settings concerned PREP at the suburban school. Grades given by teachers in other classes were found to correlate highly with daily process measures (see SAR procedure below) of student academic performance in these classes (Institute for Behavioral Research, 1972a). These grades also gave a known social rating for the skills generalized from academic and social classes (Friedman, Filipczak, and Fiordaliso, 1977). Grade comparisons significantly favoring experimental students during the first year were repeated in the second year (Filipczak, Friedman, and Reese, Note 14), and grades gained the approval of community and professional reviewers as a practical measure of short-term skill transfer.

Class grades taken one year after participation demonstrated superiority of experimental group performance and encouraged longer term, four-year follow-up (Filipczak *et al.*, 1979). Independent evaluators also found class grades to be a basic feature of practical measurement for PREP in the natural setting (Ingle and Wall, Note 15).

The rural school faculty encouraged the continued use of class grades as a measure of student skill generalization. Lack of a pure no-treatment comparison group limited the analyses possible the first year, but the non-PREP class grades of the academic treatment group significantly surpassed those of the social skills

group (Friedman *et al.*, 1977), and grades in non-PREP classes were most highly associated with PREP grades for the academic group students. During this and all remaining rural school years, PREP used no systematic daily procedure for following up on student performance in other classes, indicating that PREP training was generalizing here without additional supports.

School faculty and administration of the urban school preferred using the class grade measure, advancing the same rationale as had been given by the suburban and rural teachers. However, discussions with several urban faculty confirmed PREP staff impressions that grades in the non-PREP classes would be unlikely to reflect substantial information about the students' cognitive progress. The grades of experimental and control group students did not differ significantly over the first urban site year (Friedman *et al.*, 1977). However, during the second year, a more systematic procedure was instituted to follow up daily on student performance in other classes. Here, substantial and significant differences in favor of the experimental group were observed (Reese and Filipczak, *in press*), encouraging the use of support methods in other classes.

Generally, the use of class grades as an academic measure was encouraged by most participants in PREP, across all sites, despite questions about their reliability and validity. Their use yielded a clear discrimination between treatment groups at each site.

Follow-up Measures

The interests of high-level school officials and the funding agency spurred concern for the assessment of long-term academic maintenance with PREP students. Certain of the results and some of the difficulties encountered in gathering and interpreting follow-up data may guide similar efforts.

Table 3 indicates that the academic follow-up is partially completed, with data on reading (Gates or SAT) and school progress available

Table 3
Selected Academic Follow-up Assessments by Site and Year

Assessment	Laboratory		Suburban		Rural		Urban	
	1969-1971		1971-1972	1972-1973	1973-1974	1974-1975	1975-1976	1974-1975 1975-1976
GENERAL ACADEMIC ACHIEVEMENT:								
Data Source								
Design ^a					Planned	Planned	Planned	SAT Tests E vs. C A
Citation								A
READING ACHIEVEMENT:								
Data Source								
Design ^a								SAT Tests E vs. C A
Citation								A
SCHOOL PROGRESS:								
Data Source	Self-Report	Verified Self-Report	Verified Self-Report E vs. C C		Planned	Planned	Planned	School Records E vs. C
Design ^a	Within B							
Citation								

* Differences assessed and significance obtained in favor of treatment group(s).

^aDesign Code: Within = Within-Group Comparisons, E vs. C = Experimental and Control Comparisons, Mult = Multiple Group Comparisons
Citation Legend:

A. Filipczak *et al.*, Note 11.

B. McCombs *et al.*, 1978.

C. Wodarski and Filipczak, *in press*.

from the suburban and urban sites and limited data available from the laboratory site.

Early professional and community reviewers of PREP echoed the sentiment of the funding agency to assure some reasonable measures of student skills after treatment. However, sufficient funding was not available for an ideal follow-up that would involve naturalistic observational measures and intensive academic and social skill assessment with hard-to-trace students. In addition, severe constraints were placed on access to school records. Thus, the rationale for the follow-up became one of sampling in a number of areas, while staying within logistical and budgetary restrictions.

For academic follow-up, the measures of reading tests and progress in school, while relatively simple and inexpensive to obtain, also provided useful information for the different PREP audiences. Although the follow-up is not completed, evidence already indicates differential long-term outcome effects across groups and sites.

General academic achievement. Measures, like the SAT, were not taken at the laboratory or suburban follow-up. The reading and mathematics SAT subtests were used as pre- and post-measures at the urban site, as they permitted ready assessment of lower skilled students. Since there was evidence of positive short-term outcomes at the urban site favoring experimental groups using the SAT and other measures (Filipczak *et al.*, 1979), a follow-up on these students was deemed important, as their personal and demographic characteristics indicated that they were more "at risk" than students from other sites. Budgetary constraints, however, limited follow-up efforts to a single self-report and testing session at school and the use of school records. This follow-up was held an average of two years after the students participated in the program. As social follow-up measures were also obtained at the session, time permitted administration of only the reading subtest of the SAT.

No differences were found between the sub-

samples of experimental and control students for pre- to follow-up comparisons (Filipczak *et al.*, Note 11) on the reading subtest of the SAT. This finding did not confirm the short-term gains observed over the treatment year.

Reading achievement. Through lack of planning by the project staff, reading achievement was not part of the follow-up assessment of laboratory students.

But, as the Gates MacGinitie Reading Survey had been used in all years at the suburban site for pre-to-post assessment, it provided a useful follow-up measure. However, the time and costs of administering the test dictated that the comprehension subtest alone would be given at four-year follow-up, with only the second year's sample of students. This 43% subsample ($N = 34$) was representative of the original student groups. During this treatment year, the design consisted of five groups: academic only, academic and social, social only, family, and control. One-way analyses of variance indicated no differences among follow-up groups for the treatment year scores. However, a significant difference was found among groups at follow-up via ANOVA, and a Scheffe test identified the academic as the component performing significantly better than controls (Filipczak *et al.*, Note 11).

Though these results were from one program year, the findings parallel data reported previously which suggested that the addition of the social component to the academic component yielded no increment or decrement to academic outcomes (Filipczak, Picek, Archer, Friedman, Fiordaliso, and Lordeman, Note 16). It is also possible that logistical and political problems encountered during that particular year and described elsewhere (Filipczak and Friedman, 1978) may have masked pre-to-posttreatment differences PREP usually observed between groups.

School progress information. For each of the follow-ups, information on class grades and school year completed has been gathered. An initial four-year follow-up on the laboratory

program provided evidence of favorable treatment outcomes for overall progress in school grade completions (McCombs, Filipczak, Friedman, and Wodarski, 1978). For example, self-report information from students indicated that 53% of the follow-up group had completed high school and that two (13%) of these students were then enrolled in college. These results encouraged staff to pursue this type of information for later school groups.

A one-year follow-up of students at the suburban site indicated that experimental group students maintained higher combined mathematics and English language class grades than the control students (Wodarski, Filipczak, McCombs, Koustenis, and Rusilko, *in press*). These findings provided further impetus for a four-year follow-up, using measures such as class grades and school year completion.

However, enactment and implementation of the *Family Educational Rights and Privacy Act of 1974* helped render the school records of former suburban school participants inaccessible to PREP. Research officials of the suburban county school administration did not want to risk litigation by parents and demanded that highly formal protocols be used to secure the consent of parents in addition to the written consent documents used by PREP. Although this legislation worked to ensure privacy of school records of the primary important audience (the students themselves), it denied other audiences (the funding agency, the researchers, and juvenile program professionals) the opportunity to obtain measures of the long-term effectiveness of PREP.

In the four-year suburban follow-up, school progress information and class grades were obtained through student self-report taken in non-school settings. These reports were verified through structured interviews by independent consultants with former students' mothers or other nearest relatives.

An analysis of data from the two groups combined for followup does not reveal significant differences between groups on measures

of class grades earned since PREP. However, unreported data from PREP follow-up indicate that class grade differences favor the first year's experimental group at the four-year follow-up point.

Information was also mixed on school year completions from follow-ups with the suburban school groups. None of the individual experimental groups for either of the years had completed more school years than the controls, and, overall, the control group performed marginally better on this measure than the combined experimental groups in combined two-year analysis (Filipczak, Winett, Neale, and Rusilko, Note 17). Yet, experimental students demonstrated more academically oriented aspirations for themselves and their children, more long-term school plans, and less school absenteeism and truancy in previously unreported data.

While the academic follow-up has not demonstrated conclusive long-term effectiveness, there is some evidence for more positive short- and long-term outcomes at the suburban site. These findings suggest an important program-by-site interaction that will be further examined at the rural site.

SOCIAL MEASURES

Each program year, PREP gathered several measures of social behavior and pro-social involvement. Where possible, similar or parallel measures were obtained at follow-up.

Program Measures—Process

A rating by teachers of students' in-class behavior was used each year, but the involvement of non-PREP teachers varied across sites (Table 4). Data from the ratings were primary process measures, the bases of decisions to commence or to terminate social training.

Teacher ratings were used for the PREP laboratory school (Cohen *et al.*, 1971), where students spent mornings, attending three afternoon classes at their regular schools. Student-carried behavior rating forms, called Student

Table 4

Consistency and Change in Selected Social Program Assessments by Site and Year

Assessment	Laboratory			Suburban			Rural			Urban		
	1969-1971	1971-1972	1972-1973	1971-1972	1972-1973	1973-1974	1974-1975	1975-1976	1974-1975	1975-1976	1974-1975	1975-1976
STUDENT BEHAVIOR EVALUATION:												
Data Source	All Teachers Within A	All Teachers Within B	All Teachers Within C	All Teachers Within C	All Teachers Within C	PREP Only Within D	PREP Only Within E	PREP Only Within F	PREP Only Within G	PREP Only Within H	PREP Only Within H	PREP Only Within H
Design ^a												
Citation												
ATTENDANCE:												
Data Source	School Records Within A	School Records E vs. C I	School Records Mult* I	School Records Mult* I	School Records Mult* I	School Records Mult* J	School Program Mult* K	School Records Mult* J	School Program Mult* I	School Program Mult* I	School Program Mult* I	School Program Mult* I
Design ^a												
Citation												
DISCIPLINARY REFERRALS:												
Data Source		School Records E vs. C* I	School Records Mult* I	School Records Mult* I	School Records Mult* I	School Records Mult* J	School Records Mult* J		School Records Mult* G			
Design ^a												
Citation												
OTHER:												
Data Source	Suspension	Suspension	Suspension	Suspension	Suspension Nomination		Suspension	Ratings	Suspension	Suspension Ratings		
Design ^a		Within A	E vs. C* I	E vs. C* L	Mult* C		Mult* E	Mult* J	E vs. C* J	E vs. C* H		
Citation												

* Differences assessed and significance obtained in favor of treatment group(s).

^aDesign Code: Within = Within-Group Comparisons, E vs. C = Experimental and Control Comparisons, Mult = Multiple Group Comparisons

Citation Legend:

- A. Cohen *et al.*, 1971.
- B. Cohen *et al.*, 1974.
- C. Goding and Cohen, Note 18.
- D. Filipczak and Friedman, 1978.
- E. Filipczak and Picek, Note 19.
- F. Filipczak and Picek, Note 13.
- G. Filipczak and Friedman, Note 7.
- H. Filipczak *et al.*, Note 8.
- I. Filipczak *et al.*, *in press*.
- J. Filipczak and Wodarski, Note 12.
- K. Fiordaliso *et al.*, 1977.
- L. Wodarski and Filipczak, *in press*.

Activity Records (SAR's), were a practical means of assessing transfer of training through the students' social performance in school from simple behavior-referenced checklists. The completed forms were signed by parents and returned to the laboratory, where the data were displayed and tangible consequences were applied.

In the suburban school, staff added an academic behavior category to the SAR to explore the relationship between social behavior and academic grades (Cohen *et al.*, 1974). Although the forms were not sent home, parents still received feedback from the daily ratings, via a weekly newsletter summary. Student use of the rating form was again tied to the program's reinforcement system. In the first suburban year, the high percentage of forms returned, high average ratings, and the sharply decreased number of office referrals (Friedman, *et al.*, 1977) suggested that the teacher rating helped maintain appropriate social behavior.

In the second suburban year, the SAR was used with the "social only" and "total treatment" groups. The SAR also served as a daily measure for individual student programs—to support homework and student-parent communication. The two groups using the SAR procedure had the lowest incidence of disciplinary referrals (Goding and Cohen, Note 18), but, as other treatment variables were also in effect, it was impossible to determine the specific impact of the ratings.

At the rural site, administrators and teachers were reluctant to use the teacher ratings in non-PREP classes, fearing negative student labeling. In addition, the class scheduling used at this site and the lack of skilled aides made it impossible for non-PREP teachers to rate groups of up to 20 PREP students near or at the end of class. Nevertheless, PREP teachers saw daily behavior ratings as useful for maintaining control, using them in the PREP classes only (Filipczak and Picek, Note 19). At this site, rating data were used by project staff and participating teachers in their regularly scheduled planning and as-

essment meetings to gauge the effects of special events or disruptions at school on student performance (Picek and Filipczak, Note 4). In addition, the data were routinely reported along with academic data to the school administration and funding agency.

Urban teachers used the ratings in PREP classes only. In the first year, only a limited number of behaviors (gross behavioral categories, attendance, work completed) were cited on the form, which was designed to be completed by students and confirmed or modified by the teachers. These behaviors were problematic at the site (Filipczak and Friedman, Note 7), and the school administration sanctioned using paired tangible reinforcement to introduce students to the natural social reinforcement of school for attainment of predetermined behavior rating averages. The second year, a more multidimensional rating was implemented (Filipczak *et al.*, Note 8).

Project staff, administrators, teachers, and parents found the teacher behavior rating a useful process measure, when viewed in conjunction with academic outcomes, and adaptable to the practical and ethical concerns of those using it.

Program Measures—Outcome

PREP tried to maintain outcome measures of social change consistently over years and sites. The measures shown in Table 4 exemplify those selected to reflect the project's aims and objectives, while adhering to practical, financial, legal, and ethical constraints and differing audience needs in each setting.

School attendance. All sites used the attendance measure because of its observed relationship with delinquency (Senna, Rathus, and Siegal, 1974), because attendance could vary as a function of performance at school, and because the data could be gathered unobtrusively and easily from school records. Also, the presence of repeated significant findings in favor of treatment students across years encouraged PREP staff (Filipczak and Wodarski, Note 12).

Two factors affected the relative importance of this measure across sites. The first was the degree to which attendance was a general problem at each school, affecting its perceived importance by different audiences. Second was the presence of factors beyond the control of the project that influenced the usefulness and interpretation of the attendance measure, for example, the quality of the school records themselves. At all but one site, school records proved to be an accessible and reliable source of this measure, perhaps because federal and state funding is contingent on their completeness. School administrative policies that could inhibit or contradict the project's efforts were another influencing factor. These problems arose in both years at the urban site, where project efforts to increase attendance through shaping procedures were undercut by an administrative policy that considered a student absent unless he completed a full day of school and suspended students for class cutting.

Table 4 also shows that special interventions to improve attendance were used in three instances at two sites. In the second rural year, PREP was urged to develop a telephone- and mail-based special procedure to reduce absenteeism. This program did not treat an immediate need at that site but responded to growing interest among school administrators in practical, low-cost methods for improving attendance. The success of this procedure (Fiordaliso, Lordeman, Filipczak, and Friedman, 1977) caused PREP to attempt implementing it in the urban site, where it *was* in response to an immediate need, but where implementation and assessment proved difficult (Filipczak *et al.*, Note 8).

Disciplinary referrals. Another outcome measure selected was disciplinary referrals, because of close relationship to project goals and the ease of unobtrusive collection. PREP staff, teachers, parents, and community members all viewed disciplinary referrals as a major indicator of social adjustment in school. Table 4 indicates school records were used to obtain this measure at all but the laboratory school site,

but records were complete, reliable, and valid enough for analysis only at the suburban site. In both suburban years, significant differences on this measure were found in favor of the experimental students (Filipczak, Archer, and Friedman, *in press*). Encouraged, PREP sought to replicate these outcomes at the rural and urban sites. However, after one year at each site, use of this measure was abandoned, when project staff observed major discrepancies in record keeping policy and practice, or after the records themselves were found incomplete, inaccessible, or not subject to clear interpretation (Filipczak and Friedman, Note 7; Filipczak and Picek, Note 19).

Other social measures. Specifically, PREP also used suspensions, during six years; and teacher nominations, social checks on report cards, citizenship ratings, and retention in grade, each in one instance only. Suspension from school was another primary measure of school social adjustment interesting to most of PREP's audiences. Suspension data were gathered in the laboratory school, in both years at the suburban site, the second year at the rural site, and both years at the urban site. However, suspensions occurred with such low frequency at two sites that few or no differences could be revealed. An exception occurred in the second urban year, where differences on this measure were found significant in 8th grade and encouraging in 7th and 9th grade, all in favor of experimental students (Filipczak *et al.*, Note 8).

The other measures in this category reflect PREP's effort to seek more unique measures of social behavior, when those originally conceptualized were found to be unsuitable. One was a teacher nomination system used the second suburban year. Here, teachers were asked to construct a list of students whom they felt most needed extra social training for the next year. Significant differences were found in favor of experimental students (Goding and Cohen, Note 18). But this procedure was discontinued as its validity could be questioned, it was obtrusive, and it was time consuming to gather.

In the third rural year, a new report card form was instituted on which teachers rated students' quarterly performance for 10 social behavior categories. PREP used these naturally occurring data in analyses that yielded significant differences in favor of PREP students on 8 of the 10 measures of classroom behavior (Filipczak and Wodarski, Note 12).

The emergence of the report card measure at the rural site spurred PREP to use additional social measures from available urban school data. One of these measures was student retention in grade. On the whole, this measure showed that PREP 8th grade students were retained in grade significantly less often than controls, and favorable experimental differences were found for the 9th graders, potentially a generalization effect of the program (Filipczak and Wodarski, Note 12). A second measure, obtained unobtrusively from the urban records, was citizenship grades, assigned by homeroom teachers. Here, significant improvement was shown by 8th graders. Improvement was shown in 7th and 9th grades as well. However, comparisons were questionable here because all the experimental students from the grades were placed in the same homeroom, conducted by the PREP social skills teacher (Filipczak *et al.*, Note 8).

Follow-up Measures

The assessment of skill generalization and maintenance in PREP depended on the divergent and conflicting rationales expressed for follow-up by various audiences, such as staff, the funding agency, juvenile program officials, and evaluators. However, long-term studies remain subject principally to the availability of pertinent data sources. Under this constraint, most PREP follow-up information on social variables is in the form of student self-reports. Self-report instruments have been used in evaluations of adolescent behavioral programs (Cohen and Filipczak, 1971; Eitzen, 1975; Stump-hauzer, 1979; Tharp and Wetzell, 1969). PREP

follow-up batteries included validated scales of socialization, self-esteem, and delinquent behavior, and specific variables relevant to the program's long-term maintenance goals.

Socialization. The *Gough Socialization (SO) Scale* (Gough, 1975), a standardized measure with previously demonstrated predictive validity for discriminating between delinquents and non-delinquents (Megargee, 1972; Rosen and Schalling, 1974; Stein, Vadum, and Sarbin, 1970) was used to assess socialization. The SO scale was designed to assess social maturity and is included in the *California Psychological Inventory*. The scale was incorporated into PREP follow-up because of the presumed inverse relationship of socialization with delinquency (Hirschi, 1969; Mischel, 1973; Nye, 1973). Juvenile program officials were particularly interested in the scale's discriminative potential.

As seen in Table 5, students in the laboratory follow-up completed the scale, but comparisons with control groups were not possible (McCombs *et al.*, 1978).

The SO scale was retained for the suburban school follow-up to ensure that data gathering reflected interests of the funding agency. Although administration in both suburban years failed to reveal differences between experimental and control groups in total score, differences were apparent on several scale items. Experimental group students recorded significantly higher pro-social responses for items concerning guilt about misbehavior, hope about the future, avoidance of trouble situations, use of alcohol, and response to anti-social challenges by peers. Control students reported themselves to be from happier, closer knit families and to be less nervous when talking in public (Filipczak *et al.*, Note 17).

The lack of significant differences on the total scale at follow-up suggested that experimental and control students did not differ in their ability to take pro-social roles. The time involved in administration and the observed lack of concurrent validity with more concrete social measures led PREP staff to omit all but a few

Table 5
Selected Social Follow-up Assessments by Site and Year

Assessment	Laboratory			Suburban			Rural			Urban		
	1969-1971	1971-1972	1972-1973	1973-1974	1974-1975	1975-1976	1974-1975	1975-1976	1974-1975	1975-1976	1974-1975	1975-1976
SOCIALIZATION:												
Data Source	Complete Gough Within A	Complete Gough E vs. C* B	Complete Gough Mult* C	Planned	Planned	Planned	Planned	Planned	Partial Gough E vs. C	Partial Gough E vs. C	Partial Gough E vs. C	Partial Gough E vs. C
Design ^a												
Citation	A	B	C									
SELF-ESTEEM:												
Data Source	Rosenberg Within A	Rosenberg E vs. C* B	Rosenberg Mult* D	Planned	Planned	Planned	Planned	Planned	Rosenberg E vs. C	Rosenberg E vs. C	Rosenberg E vs. C	Rosenberg E vs. C
Design ^a												
Citation	A	B	D									
INCIDENCE OF DELINQUENT BEHAVIOR:												
Data Source	Court Record Within A	Verified Self-Report E vs. C* B	Verified Self-Report Mult* D	Planned	Planned	Planned	Planned	Planned	Verified Self-Report E vs. C	Verified Self-Report E vs. C	Verified Self-Report E vs. C	Verified Self-Report E vs. C
Design ^a												
Citation	A	B	D									
ATTENDANCE:												
Data Source	Self-Report Within	Verified Self-Report E vs. C B	Verified Self-Report E vs. C	Planned	Planned	Planned	Planned	Planned	School Record E vs. C	School Record E vs. C	School Record E vs. C	School Record E vs. C
Design ^a												
Citation												
EMPLOYMENT:												
Data Source	Employer Interview Within	Verified Self-Report E vs. C B	Verified Self-Report E vs. C	Planned	Planned	Planned	Planned	Planned	Verified Self-Report E vs. C	Verified Self-Report E vs. C	Verified Self-Report E vs. C	Verified Self-Report E vs. C
Design ^a												
Citation												

* Differences assessed and significance obtained in favor of treatment group(s).

^aDesign Code: Within = Within-Group Comparisons, E vs. C = Experimental and Control Comparisons, Mult = Multiple Group Comparisons
Citation Legend:

A. McCombs *et al.*, 1978.

C. McCombs *et al.*, Note 20.

B. Wodarski and Filipczak, *in press*.

D. Filipczak *et al.*, Note 17.

items (those which provided discrimination among groups) as part of the urban follow-up.

Self-esteem. To determine the long-term effects of PREP on student attitudes self-esteem was assessed, reflecting audience concern for measures extending beyond behavioral conceptions. PREP included the *Rosenberg Self-Esteem Scale* (Rosenberg, 1965) throughout the follow-up. The Rosenberg is a 10-item Guttman scale, proven reproducible and scalable, and is designed to assess the extent to which an individual rates his self-esteem in reference to others. Students who participated in the laboratory program indicated a high level of personal satisfaction at follow-up (McCombs *et al.*, 1978), but, once again, no comparisons could be made.

First year suburban school students also completed the scale at follow-up (Wodarski and Filipczak, *in press*), and *t* tests revealed no significant differences between groups in total scale score. Second year participants in the academic treatment group scored significantly higher on the total scale than other treatment or control groups (McCombs, Filipczak, Rusilko, Koustenis, and Wodarski, Note 20). This added support to the conception that academic training yields benefits in social areas.

Delinquent behavior. This assessment proved difficult but was required to satisfy both project and funding agency staff that substantial knowledge was obtained of student status and criminal offenses. The inaccessibility of school and court records (Filipczak *et al.*, Note 1) most hampered the program's attempt to establish the impact of PREP on students' formally recorded delinquent behavior. Verifiable self-report instruments were ultimately used in place of most police and court record data (Table 5).

The only instance in which direct court records were available to PREP staff involved the students in the laboratory setting (McCombs *et al.*, 1978). Here, records indicated that the frequency of student arrests and detentions declined significantly during the program, maintained at this low level for a year after treat-

ment, but returned to the pre-program level by the four-year follow-up.

Subsequent follow-up efforts asked students to self-report involvement with police, court, probation, and corrections institutions because court records could not be obtained. Students also completed an expanded version of the *Nye and Short Self-Report of Delinquency Scale* (Nye, 1973; Nye and Short, 1957), which included frequency ratings of felony and misdemeanor violations, such as auto theft, petty theft, and vandalism. Students reported acts committed since PREP involvement.

No significant differences on total scores were found between treatment and nontreatment groups for either year of the suburban follow-up. However, first year experimental students reported significantly less involvement in gang fights than controls (Wodarski and Filipczak, *in press*). Student self-reports were verified by parent responses for 20% of follow-up items, yielding agreement of 75% for these comparable items.

Although no significant differences were observed on specific delinquent behavior measures, items from the *Gough Socialization Scale* did show significant and relevant attitude differences at follow-up (Filipczak, *et al.*, Note 17). Experimental group students, particularly participants in academic treatment, indicated less willingness to act on dares, less excessive use of alcohol, less trouble seeking, and a more prosocial response to what others would say about their involvement in delinquent acts.

School attendance. Project staff, school administrators, and evaluators saw attendance as a critical measure of both short- and long-term outcome and as a necessary prerequisite to effective in-school treatment of students. During intervention periods, school attendance records were relatively easy to obtain or, where inaccurate, easy for project staff to maintain. The use of such measures for PREP follow-up depended on access to individual school records. Table 5 shows that, when access was denied by

administrators, student self-report became the measure of choice, particularly to the extent that it could be verified (Filipczak *et al.*, Note 1).

As laboratory students were well over compulsory school age at follow-up, they were not asked to provide retrospective reports on attendance. One-year follow-up for the first suburban year indicated that experimental group attendance was higher than for controls (Wodarski *et al.*, *in press*). At four-year follow-up, suburban students were asked to check categorical levels of attendance for the school year just completed, but student reports of attendance for both years at the four-year follow-up point revealed no significant differences between groups.

Because follow-up with urban students took place only two to three years after program participation, most participants were still of school age. A review of post-program attendance records at these schools was undertaken to provide verification of student self-report. These data are not completely analyzed.

Despite the face validity of attendance as an important measure of long-term success, reliable data were difficult to obtain, and substantial differences between groups have not been observed.

Employment. Status of the students' employment represented a major concern of community reviewers, project staff, and the funding agency. At the four-year follow-up, few students of the laboratory program endorsed access to their employer as a data source, indicating that seeking such information could potentially embarrass or harm them. Over two-thirds of PREP laboratory students who provided follow-up self-report indicated that they had been employed in the four years since leaving the program (McCombs *et al.*, 1978). Half of the students were still employed, while the rest were not (two attending college), suggesting difficulties in maintaining, rather than securing, employment.

Privacy issues in the suburban school system (Filipczak *et al.*, Note 1) limited PREP staff to

self-report data for this variable at follow-up, with parents verifying this information. Four-year follow-up of suburban students revealed a similar pattern in terms of employment and available data for both years, with no significant differences between groups. Here, however, urban follow-up data (as yet unpublished) show that PREP students in the second program year were more likely to be working than their control counterparts.

Again, the problems encountered in getting valid information from reliable sources limited assessment on a measure of demonstrable importance to the long-term success of former PREP students.

SUMMARY AND CONCLUSIONS

The multivariate assessment described here represents a small fraction of the procedures recommended for use by project staff and relevant audiences and implemented over PREP's term. Attempts at measurement consistency or adherence to highly observable methods became incompatible with the need to respond to changing audiences, audience priorities, and program outcome information. PREP appears to exemplify the conditions Azrin (1977) envisioned in noting that the proliferation of applied behavior interventions will demand implementation of wider ranging and more diverse assessment procedures.

The selection of appropriate measurement instruments and procedures presents few problems in basic research settings where traditional concerns for reliability and validity are paramount. However, for survival, this community-based applied research program was forced to transcend these basic criteria and choose procedures appropriate to both the stated objectives and to the potential unintended effects. PREP found several of its audiences placed increasing emphasis on the generalizability of treatments, particularly the extent to which learned skills maintain over time.

The overall context within which applied community-based research programs operate enforces other demands on its execution. The treatment efforts and program outcomes of PREP were sharply influenced through standards imposed by the local community in cooperative goal setting or by community responses to project events. In certain instances, intense community reactions to ethical and legal issues in youth treatment forced PREP to redirect or abandon procedures judged by independent review boards to be important for adequate assessment and non-injurious to participants. Community and teacher insistence on the maintenance of behavior ratings in each school showed unexpected support for use of observable methods. Similarly, in later years and settings, the needs of teachers yielded practical changes and reductions in the earlier and more complex student assignment completion measures.

Management policies and procedures of the host school systems also limited project operations. Unreliable data collection methods or research policies that inhibited access to school data at follow-up created unforeseen bars to completion of planned activities. Yet, school administrators and faculty supported PREP during times of intense community pressure and promoted good relations with non-PREP teachers that permitted gathering daily data on generalization of student behavior.

Other more persuasive contextual problems confounded PREP's intended research and demonstration. For example, both the staff time and financial requirements for mounting systematic extension of during and posttreatment assessment seriously exceeded the resources of the project and its funding agency. This yielded a decrement in the intended program that was a reflection of practical intrusions into theoretical plans.

Yet, the PREP experience does not seem unique among community-based applied research programs. Departure from the controlled research setting compels the behavioral investi-

gator to observe those requirements of the environment that routinely confront community service agencies. These departures from traditional research concerns prompt an imaginative translation of project objectives to suit community needs and a resourceful integration of community demands into effective, valid, and reliable assessment procedures.

Future large-scale applied research programs may face a dual rationale for multivariate assessment. First, their staff will demand that each of the *intended* program effects is assessed as rigorously as has been the case in the past; and, as the scope and dimensions of the objectives for the programs increase, so will the scope of assessment. Second, as behavioral training procedures become better known and more widely applied in service settings, persons outside the research staff can systematically and knowledgeably challenge limited assessment for complex situations. The outsiders can force the researcher to assess the occurrence and level of outcomes *unintended* by the researcher.

The survival of large-scale applied research studies will undoubtedly continue to depend on the *quality* of the research being conducted. As such studies are more frequently undertaken, however, social context, social values, and social policy, reflected in the demands by different audiences for different assessments, may dictate both the direction of treatment programs and their evaluation.

REFERENCE NOTES

1. Filipczak, J., McCombs, D., Rusilko, S., and Koustenis, G. *Follow-up on juvenile behavior intervention: Progress and impediments in execution*. Paper presented at the meeting of the American Psychological Association, San Francisco, September 1977.
2. Cohen, H. L., Filipczak, J., Slavin, J., and Green, D. *Project interim report: The PICA Project, year 2*. Silver Spring, Maryland: Institute for Behavioral Research, 1970.
3. Friedman, R. M., Filipczak, J., Fiordaliso, R., and Picek, E. J. *Standardized test performance to evaluate a behavioral program in the schools?* Paper presented at the meeting of the American Psychological Association, Chicago, August 1975.

4. Picek, E. J. and Filipczak, J. *First year interim report: The PREP Project—Rural Middle School*. Silver Spring, Maryland: Institute for Behavioral Research, 1974.
5. Fiordaliso, R. and Archer, M. *Generalization of academic performance across classroom settings*. Paper presented at the meeting of the Association for Advancement of Behavior Therapy, Atlanta, December 1977.
6. Fiordaliso, R., Friedman, R. M., Brotman, G., and Lordeman, A. *Effects of contingent versus non-contingent reinforcement on high and low base rate behaviors*. Paper presented at the meeting of the Association for Advancement of Behavior Therapy, New York, December 1976.
7. Filipczak, J. and Friedman, R. M. *The PREP Project, Urban Junior High School: Progress report, August 1975*. Silver Spring, Maryland: Institute for Behavioral Research, 1975.
8. Filipczak, J., Reese, S., Fennell, S., Bass, G., Kent, T., and Gilmore, S. *Progress report: The IMPACT Program, Urban Junior High School*. Silver Spring, Maryland: Institute for Behavioral Research, 1976.
9. Filipczak, J. *Practical problems in translating in-school research to community use*. Paper presented at the meeting of the Association for Advancement of Behavior Therapy, Miami Beach, December 1973.
10. Friedman, R. M., Archer, M., and Lordeman, A. *Five years in public schools: A summary of PREP findings*. Paper presented at the meeting of the American Psychological Association, Washington, D.C., September 1976.
11. Filipczak, J., Neale, M. S., Winett, R. A., and Rusilko, S. *Differential effects of behavior intervention in school: Four-year follow-up*. Paper presented at the meeting of the American Educational Research Association, San Francisco, April 1979.
12. Filipczak, J. and Wodarski, J. S. *Behavioral intervention in public schools: Implementing and evaluating a model*. Manuscript submitted for publication, 1979.
13. Filipczak, J., Friedman, R. M., Fiordaliso, R., and Picek, E. J. *Relationships between standardized test performance and student academic and social behavior measures*. Paper presented at the meeting of the American Psychological Association, Chicago, 1975.
14. Filipczak, J., Friedman, R. M., and Reese, S. C. *School-based intervention with troublesome juveniles: Who needs it?* Paper presented at the meeting of the American Psychological Association, Washington, D.C., September 1976.
15. Ingle, R. B. and Wall, R. *A review and assessment of Preparation through Responsive Educational Programs, PREP*. Silver Spring, Maryland: Institute for Behavioral Research, 1978.
16. Filipczak, J., Picek, E. J., Archer, M., Friedman, R. M., Fiordaliso, R., and Lordeman, A. *Third year interim report: The PREP Project, Rural Middle School*. Silver Spring, Maryland: Institute for Behavioral Research, 1976.
17. Filipczak, J., Winett, R. A., Neale, M. S., and Rusilko, S. *Follow-up with disruptive adolescents: Social validation and outcomes*. Paper presented at the meeting of the American Psychological Association, Toronto, 1978.
18. Goding, I. D. and Cohen, H. L. *Teaching interpersonal skills through regular classroom procedures: An assessment of the effectiveness*. Paper presented at the meeting of the American Educational Research Association, Chicago, April 1974.
19. Filipczak, J. and Picek, E. J. *The PREP Project, Rural Middle School: Second Year Report*. Silver Spring, Maryland: Institute for Behavioral Research, 1975.
20. McCombs, D., Filipczak, J., Rusilko, S., Koustenis, G., and Wodarski, J. *Follow-up on behavior development with disruptive juveniles in a public school*. Paper presented at the meeting of the Association for Advancement of Behavior Therapy, Atlanta, December 1977.

REFERENCES

- Azrin, N. H. A strategy for applied research: Learning based but outcome oriented. *American Psychologist*, 1977, **32**, 140-149.
- Cohen, H. L. and Filipczak, J. *A new learning environment*. San Francisco: Jossey-Bass, 1971.
- Cohen, H. L., Filipczak, J., Boren, J., Goding, I., Storm, R., Bishop, R. M., and Breiling, J. *Academic and social behavior change in a public school setting*. Silver Spring, Md.: Educational Facility Press, Institute for Behavioral Research, 1974.
- Cohen, H. L., Filipczak, J., Slavin, J., and Boren, J. *Programming Interpersonal Curricula for Adolescents (PICA)—Project year three: A laboratory model*. Silver Spring, Md.: Educational Facility Press, Institute for Behavioral Research, 1971.
- Dietz, S. M. Current status of applied behavior analysis: Science versus technology. *American Psychologist*, 1978, **33**, 805-814.
- Eitzen, D. S. The effects of behavior modification on the attitudes of delinquents. *Behaviour Research and Therapy*, 1975, **13**, 295-299.
- Filipczak, J., Archer, M., and Friedman, R. M. In-school social skills training with disruptive adolescents. *Behavior Modification*, in press.
- Filipczak, J., Breiling, J., and Storm, R. Programming for disruptive and low-achieving students: An experimental in-school alternative. *Journal of the International Association of Pupil Personnel Workers*, 1973, **17**, 138-142.

- Filipczak, J. and Friedman, R. M. Some controls on applied research in the public secondary school. In T. A. Brigham and A. C. Catania (Eds), *Handbook of applied behavior research: Social and instructional processes*. New York: Irvington Press/Halstead Press, 1978.
- Filipczak, J., Friedman, R. M., and Reese, S. C. PREP: Educational programming to prevent juvenile problems. In J. S. Stumphauzer (Ed), *Progress in behavior therapy with delinquents*. Springfield, Ill.: Charles C Thomas, 1979.
- Fiordaliso, R., Lordeman, A., Filipczak, J., and Friedman, R. M. Effects of feedback on absenteeism in the junior high school. *Journal of Educational Research*, 1977, **70**, 188-192.
- Friedman, R. M., Filipczak, J., and Fiordaliso, R. Within-school generalization of the PREP academic program. *Behavior Therapy*, 1977, **8**, 986-995.
- Gough, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Cal.: Consulting Psychologists Press, 1975.
- Heaton, R. C., Safer, D. J., Allen, R. P., Spinnato, N. C., and Prumo, F. M. A motivational environment for behaviorally defiant junior high school students. *Journal of Abnormal Child Psychology*, 1976, **4**, 263-275.
- Hirschi, T. *Causes of delinquency*. Berkeley: University of California Press, 1969.
- Institute for Behavioral Research. *PREP film program supplement*. Silver Spring, Md.: Author, 1972. (a)
- Institute for Behavioral Research (Producer). *PREP: Preparation through Responsive Educational Programs*. Silver Spring, Md.: Producer, 1972. (Film) (b)
- McCombs, D., Filipczak, J., Friedman, R. M., and Wodarski, J. S. Long-term follow-up of behavior modification with high-risk adolescents. *Criminal Justice and Behavior*, 1978, **5**(1), 21-34.
- McKee, J. M. and Clements, C. A behavioral approach to learning: The Draper model. In H. Rickard (Ed), *Behavioral intervention in human problems*. New York: Pergamon Press, 1971.
- Megargee, E. I. *The California Psychological Inventory handbook*. San Francisco: Jossey-Bass, 1972.
- Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, **80**, 252-383.
- Nye, I. *Family relationships and delinquent behavior*. New York: Wiley, 1973.
- Nye, F. and Short, J. Scaling delinquent behavior. *American Sociological Review*, 1957, **22**, 326-331.
- Reese, S. C. and Filipczak, J. Generalization across setting, behavior, and time in an educational setting. *Behavior Modification*, in press.
- Rosen, A. and Schalling, D. On the validity of the California Psychological Inventory socialization scale: A multi-variate approach. *Journal of Consulting and Clinical Psychology*, 1974, **42**, 757-765.
- Rosenberg, M. *Society and the adolescent self-image*. Princeton, N.J.: Princeton University Press, 1965.
- Senna, J., Rathus, S. A., and Siegal, L. Delinquent behavior and academic investment among suburban youth. *Adolescence*, 1974, **9**, 481-494.
- Stein, K. B., Vadum, A. C., and Sarbin, T. R. Socialization and delinquency: A study of false negatives and false positives in prediction. *The Psychological Record*, 1970, **20**, 353-364.
- Stumphauzer, J. S. (Ed) *Progress in behavior therapy with delinquents*. Springfield, Ill.: Charles C Thomas, 1979.
- Tharp, R. and Wetzel, R. *Behavior modification in the natural environment*. New York: Academic Press, 1969.
- Weinberg, S. G. and Breiling, J. Limits to comparative analysis. *CEDR Quarterly*, 1977, **10**(3), 10-13.
- Wodarski, J. S. and Filipczak, J. Long-term follow-up of behavior modification with high-risk adolescents: Cohort Two. In R. Allen, D. Safer, and R. Heaton (Eds), *Alternative programs for disruptive youths in secondary schools*. Baltimore, Md.: University Park Press, in press.
- Wodarski, J. S., Filipczak, J. A., McCombs, D., Koustenis, G., and Rusilko, S. Followup on behavioral intervention with troublesome adolescents. *Journal of Behavior Therapy and Experimental Psychiatry*, in press.
- Wolf, M. M. Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 1978, **11**, 203-214.
- Yates, A. J. Research methods in behavior modification: A comparative evaluation. In M. Hersen, R. Eisler, and P. Miller (Eds), *Progress in behavior modification* (Vol. 2). New York: Academic Press, 1976.

Received 10 October 1978.

(Final Acceptance 24 July 1979.)