

# APPENDICES

## APPENDIX

### A

#### Standardized Data Collection for Large-Scale Program Evaluation: An Assessment of the YEDPA-SAS Experience

*Charles F. Turner*

The Youth Employment and Demonstration Projects Act (YEDPA), as noted in Chapter 3, provided the Department of Labor (DOL) and its new Office of Youth Programs (OYP) with a mandate to test the relative efficacy of different methods of dealing with the employment problems of young Americans. The legislative concern with learning "what works for whom" was consistent with the frequently stated contention that decades of federal funding for similar programs had not yielded much in the way of reliable knowledge. And so, a key element of YEDPA's knowledge development strategy was the establishment of a standardized system for the systematic collection of data on the progress of program participants and the services provided by YEDPA programs.

#### STANDARDIZED ASSESSMENT SYSTEM

In order "to document administrative outcomes, to monitor performance, and to continually assess program impacts and lessons" from YEDPA programs, the Office of Youth Programs launched a large-scale data gathering operation in collaboration with the Educational Testing Service (ETS). The intent of the data gathering was to develop a standardized data base with which the performance of the various programs that YEDPA comprised could be assessed. This data gathering plan, called the Standardized Assessment System (SAS), was ambitious in its aim. SAS was intended to provide preprogram, postprogram, and follow-up data (3 and 8 months after program completion) for almost 50 percent of the youth served by these programs (Taggart, 1980).

The SAS data base is an important component of the YEDPA knowledge development enterprise not only because it was a salient feature of the YEDPA effort, but also because it provided the basic data used in evaluating a large number of the YEDPA programs. The characteristics

---

Charles F. Turner was senior research associate with the committee.

of this data base are thus of concern to us in evaluating what was learned from the YEDPA experience. In the following pages we describe the SAS data collection procedures and evaluate the characteristics of the data obtained, e.g., the coverage of the sample and the reliability and validity of the measurements.

### Data Collection Instruments

The SAS data collection instruments included an intake interview, called the Individual Participant Profile (IPP); a reading test (STEP); a battery of seven measures of occupational knowledge, attitudes, and related skills administered preprogram and postprogram; a program completion interview; interviews at 3 and 8 months postprogram; and evaluations by counselors (postprogram) and employers or work supervisors (postprogram and 3 and 8 months later). In addition, data were collected from program sites concerning the implementation of the program and the services offered, and data were also collected from "control" groups recruited by program operators to provide comparison samples for program evaluation.

In this section each of the data collection instruments is briefly described. The descriptions of the instruments are taken from The Standardized Assessment System for Youth Demonstration Projects (Educational Testing Service, 1980). Where suitable we have used the ETS phrasing or paraphrased the descriptions without repeated citation of the source.

#### Individual Participant Profile

The Individual Participant Profile was used to record information on 49 participant characteristics as well as status while in the program and at termination. These data essentially duplicated the standard information gathered on each participant in all Comprehensive Employment and Training Act (CETA) programs. The first 29 items were largely demographic, covering such information as the individual's age, sex, race, and economic, educational, and labor-force status--all at time of entry into the youth program. The remaining 20 items were "program status" items, which indicated the status of the participant at the time of program completion or termination. These included such information as entry and termination dates, total hours spent in the program, whether the program provided the participant with academic credit, and specific forms of "positive" and "nonpositive" termination. (A set of definitions accompanying the IPP form defined each item in some detail and how it was to be completed by the youth program project personnel from their project records.)

#### STEP Reading Scale

The STEP reading scale was a short (10 to 15 minutes) measure of reading skill that was intended to cover the wide range of reading

levels found among the YEDPA enrollees (approximately fourth to ninth grade reading level by ETS's estimate). Twenty items were selected from the STEP locator tests covering fourth to ninth grade reading levels. Those locator tests are short reading-comprehension measures ordinarily used as screening devices for deciding which level of the full STEP achievement tests is suitable for administration.

#### Job Knowledge and Attitudes Battery

Measures chosen for incorporation in the Job Knowledge and Attitudes battery were intended to reflect YEDPA program objectives while still being compatible with the characteristics of the trainee population and the operational constraints of the youth projects. As a starting point, five behavioral areas thought to be affected by YEDPA program participation were defined by the Office of Youth Programs. These were considered to encompass the objectives of a vast majority of the YEDPA projects and were designated as (1) career decision making, awareness, and capability, (2) self-image, (3) work attitudes, (4) job search capability, and (5) occupational sex stereotyping.

Criticism of the design and administration of conventional paper-and-pencil tests used with similar youth led SAS designers to seek measures that were relatively short, presented orally, pictorial as well as verbal, and appropriate in level and style of language for adolescents or young adults of low reading skill. In addition the battery allowed the item responses to be marked directly in the test booklet. Examples of items from each of the Job Knowledge and Attitudes battery are shown in Figure A.1.

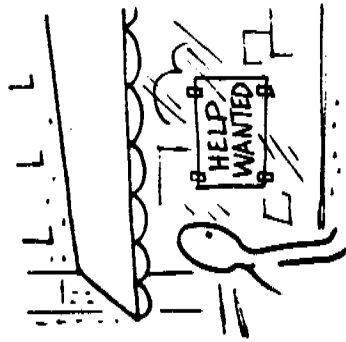
The designers of SAS chose two measures to assess what they termed career decision making, awareness, and capability performance. One measure dealt with the "vocational maturity" of adolescents in making appropriate career decisions, and the other with the youth's knowledge of what is required for carrying out different jobs.

Vocational Attitude Scale This scale contained 30 verbal items, which were scorable as three 10-item subscales. Those scales were designated as "Decisiveness," "Involvement," and "Independence" in career decision making. The respondent indicated his or her agreement or disagreement with each of 30 statements about vocational careers and employment.

Job Knowledge Test This 33-item scale dealt with the qualifications, requirements, and tasks involved in various jobs. The items, in multiple-choice format, required the respondent to indicate the correct response to questions about the specific occupations depicted.

Self-Esteem Scale Youth programs often seek to enhance the participant's feelings of personal value, or self-worth, with the expectation that improved self-perception will stimulate more success-oriented social and vocational adjustment behaviors. The SAS

VOCATIONAL ATTITUDES



How would you feel if you saw this sign?

- I might try for a job in that store, but they probably wouldn't want me.
- They would turn me down cold and wouldn't think I'm worth giving a chance.
- They would think I'm worth hiring for a job in that store.

SEX ROLE STEREOTYPING

Fire fighters work at putting out fires.

Who do you think SHOULD be fire fighters?

- Only women
- More women than men
- About the same number of men and women
- More men than women
- Only men

WORK RELATED ATTITUDES

Most bosses have it in for you and give you a hard time.

- strongly agree
- somewhat agree
- somewhat disagree
- strongly disagree

JOB SEARCH SKILLS

**SALES HELP WANTED**

SALESPERSON  
CAMERAS  
APPLIANCES  
RADIO & TV

**ROBERTSON'S  
VALLEY STREAM**

Top notch experienced sales person needed for our newest store in the Green Acres Shopping Center.

1 PM - 10 PM, 5 days

or

8 PM - 10 PM

High Salary + PMs

Opportunity to Advance  
Excellent Benefit Program

Apply at:

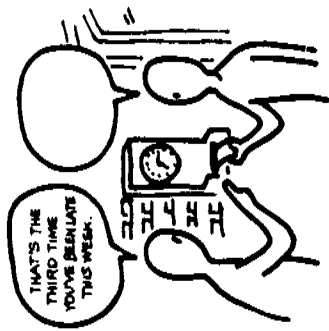
Employment Office  
480 Sunrise Hwy., Valley Stream  
An Equal Opportunity Employer

What would you do on this job?

- Fix cameras and radios.
- Sell things in a store.
- Teach people how to drive a car.
- Sell houses to people.

JOB HOLDING SKILLS

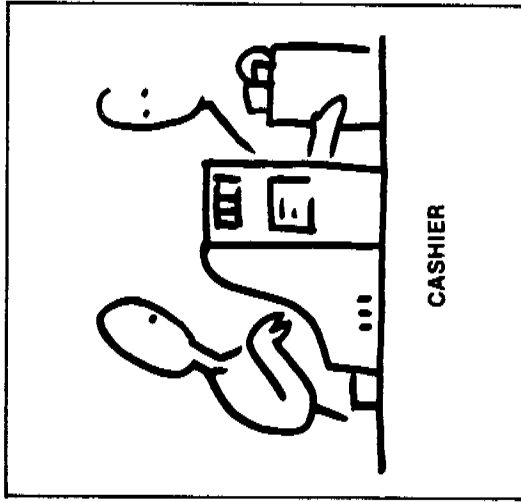
SELF ESTEEM



- What would you say to a boss who said this to you?
- Sorry, I'll try to get here on time from now on.
  - It's tough for me to make it here on time but I'll try.
  - I do my work—what's the difference if I'm late.

- T  F
- It's probably just as easy to be successful in one occupation as it is in another.
- T  F
- It doesn't matter which job you choose as long as it pays well.

JOB KNOWLEDGE



- Where would you work on this job most of the time
- In a factory on the assembly line
  - In the kitchen of a restaurant
  - In a store
  - In your own home

FIGURE A.1 Examples of items from Job Knowledge and Attitudes battery.  
 SOURCE: Educational Testing Service (1980).

designers included one measure that attempted to define the level at which the program participant rated his or her personal value. The self-esteem scale was a 15-item scale containing pictorial and verbal material used to assess perceived self-worth in terms of expectations for acceptance or achievement in various social, vocational, and educational settings. The respondent indicated, on a three-point scale, the degree to which he or she would be successful or receive acceptance in the specific situation portrayed.

Work-Related Attitudes Inventory This inventory was intended to measure the youth's views about jobs, the importance of working, appropriate ways of behaving in job settings, and general feelings about his or her capabilities for succeeding in a work situation. The inventory contained 16 items that provided both a total score and scores for three subscales defined as "Optimism," "Self-Confidence," and "Unsocialized Attitudes." The response to each of the attitudinal statements was based on a four-point scale of degree of agreement with, or applicability of, the statement.

Job Holding Skills Scale This scale dealt with respondent awareness of appropriate on-the-job behaviors in situations involving interaction with supervisors and coworkers. This 11-item scale, containing pictorial and verbal material, required the respondent to indicate which one of three alternatives best defined what his or her response would be in the situation described. (Response alternatives were scaled in terms of "most" to "least" acceptable behaviors for maintaining employment.)

Job Seeking Skills Test This test was intended to measure elementary skills essential for undertaking an employment search. This test had 17 items that sampled some of the skills needed to initiate an employment search, interpret information about prospective jobs (in newspaper want ads), and understand the information requirements for filling out a job application. The items, in a multiple-choice format, required selection of the one correct response to each question.

Sex Stereotyping of Adult Occupations Scale This scale attempted to measure attitudinal perceptions of sex roles in occupational choice. This relatively short (21 item) verbal scale presented job titles along with a one-sentence description of each job and required the respondent to indicate "who should be a \_\_\_\_\_" (job title as given). A five-point response scale ranged from "only women" to "only men."

#### Project and Process Data

In addition to the range of information collected on program participants and controls, the SAS attempted to measure the types of

activities, the progress of program implementation, and the range of services being offered at each program site. This information was expected to be of potential use not only as contextual data for the analysis of program outcomes, but also as data for reports to managers and policymakers about the implementation of the various YEDPA programs.

The Project and Process Information questionnaire contained six sets of questions that reported on key site-specific variables in quantitative terms. First, basic information was gathered about the site, setting, and sponsors of the project. Second, the project was described in terms of its services, activities, and goals. Third, the linkages involved in the project were described. Fourth, the staff involved in the project were profiled. Fifth, the project stability and the position of the project on the learning curve were assessed. Finally, the project costs were measured.

#### Outcome Measures

The outcomes of the programs were measured at program completion and 3 and 8 months after program departure. Two questionnaires were used for this purpose: the "Program Completion Survey" and the "Program Follow-up Survey." (The same instrument was used 3 and 8 months postprogram.)

Program Completion Survey This questionnaire contained 48 items, most of which were phrased as questions to be presented to the youth at the time he or she had completed or was leaving the training program. They covered the participant's activities in the program, attitudes about the program, job and educational aspirations, and expectations and social-community adjustments. The questions were intended for oral presentation to the individual by an interviewer. (A parallel questionnaire containing similar material was designed for use with control group members and was designated the "Control Group Status Survey.")

Program Follow-up Survey This 50-item questionnaire was designed to be administered orally to the individual by an interviewer, who also recorded the participant's responses. The survey was intended for use 3 months after the participant had left the training program and again at 8 months following program participation. Questions dealt with the former participant's posttraining experiences in areas of employment, education, social adjustments, and future plans. (A parallel version of the follow-up survey was used with control group members and was designated the "Control Group Follow-up Survey.") In addition, a five-item Employer Rating Form was to be completed by the present (or most recent) employer. (Permission to interview the employer had to be granted by the youth.)

### Concerns about Instrument Reliability and Validity

In introducing the SAS measuring instruments, the designers at the Educational Testing Service warned that (Educational Testing service, 1980)

more careful testing of the instruments would have been preferable but it was necessary to develop these measures while implementing certain programs. The instruments . . . represent the best possible compromise between the many constraints at the time the system was implemented.

A particular concern expressed by the SAS designers involved the nature of the youth population from whom data were being collected. Given a population characterized as economically disadvantaged and largely products of inner-city school systems, they anticipated that the validity of any available paper-and-pencil test might be suspect. For this reason the documentation of the SAS instruments stressed the (1) use of measures that employ pictures as well as words, (2) use of an administrator who would read items aloud so that the youth could follow along, and (3) the administration of the tests to small groups--so that literacy (or other) problems might be more easily detected.

Despite these precautions, it can never be assured in a data gathering operation such as SAS that measurements were made in the manner prescribed. The test administrators were not ETS employees, but rather program personnel assigned to fulfill YEDPA's "data reporting" requirement. While ETS did provide instruction to one person at each program site, that person was not necessarily the one who administered the measurements. Moreover, staff turnover may have put some people in the position of serving as test administrator with little or no (or wrong) instruction on how to administer the instruments. Since one of the canons of testing is that the manner of test administration can have important effects on measurement, it is natural that concerns about the reliability and validity of the SAS measurements were voiced by outsiders--as well as by ETS.

Almost all of the SAS scales used previously published tests, and there did exist a literature that documented the characteristics of the scales and estimated their reliability and predictive validity with various populations. These populations, however, were not identical to the YEDPA youth who would be tested with the SAS. Thus, it did not necessarily follow that the readings of test reliability and validity obtained from these groups could be generalized to the youth population targeted by YEDPA.

In its 1980 report on the Standardized Assessment System, ETS presented evidence for the reliability and validity of the SAS scales.<sup>1</sup> Some of this evidence predates YEDPA and may have been used

---

<sup>1</sup>ETS (1980) presents estimates of reliability and validity in cases where there are "significant" results (p less than .01 or p less than



in the decision making about which instruments to use in SAS. The evidence is derived from studies of small samples of youths participating in Neighborhood Youth Corps (NYC) and Opportunities Industrialization Center (OIC) training programs. For four of the SAS scales, Table A.1 presents the correlations found between scale scores and various criteria of "success" in these programs. Reported correlations range from .18 to .36. Two measures show significant correlations with success in finding employment after program completion--the Job Knowledge scale ( $r = .22$  in NYC sample) and the Job Search Skills scale ( $r = .36$  in NYC sample, and  $.21$  in OIC sample). The other two scales, Job Holding and Self-Esteem, do not show significant associations with postprogram employment, but do show positive associations with evaluations given by guidance counselors and work training supervisors.

The 1980 report on SAS also provides early SAS data from samples of high school seniors participating in the Youth Career Development project ( $n = 1,666$ ) and their control group ( $n = 1,590$ ). Estimates of predictive validity using selected criterion measures (and Cronbach's alpha for the scales) are shown in Table A.2. The range of correlations for this sample are generally lower than those found in the earlier studies. In particular, only two scales (Vocational Attitudes and Work-related Attitudes) show significant correlations with postprogram activity (coded 2 for full-time school or work, 1 for part-time school or work, and 0 otherwise). These correlations were very modest in size ( $r = .12$  and  $.10$ ). The scales did show somewhat higher correlations with level of present job and a negative correlation with amount of time required to find the present job.

Overall, however, the preliminary evidence presented by ETS suggests that (1) the seven scales are not powerful predictors of postprogram employment and (2) the measurement characteristics of these scales when administered in SAS may be different from those found elsewhere. (Whether the latter might be a function of the population tested, lack of standardization in administration, or some other cause, is difficult to say.)

---

.05). Thus it is not possible in Table A.1 and A.2 to report their estimates for all variables and for each criterion measure.

In selecting ETS "validity" measures to reproduce in Table A.2 and in designing our own analyses (reported in Table A.10 and A.12) we have focused on the prediction of future rather than concurrent outcomes where the outcome variables involved assessments by observers other than the subject (e.g., an employer's evaluation of the subject at 3 months postprogram) or involved reports of relatively objective statuses (e.g., Are you employed full time?). We believe that this procedure provides more appropriate information about the usefulness (for program evaluation) of the SAS assessment battery than procedures that depend exclusively on more subjective reports from the respondent (e.g., assessments of job satisfaction or adjustment).

TABLE A.1 ETS Estimates of Predictive Validity of SAS Attitude and Knowledge Measurements

SAS Measurement	Criterion Predicted	Sample (n)	r
Job knowledge	Work supervisor rating	NYC (109)	.32
	Counselor rating	NYC (109)	.25
	Counselor rating	OIC (220)	.19
	Vocational skills instructor rating	OIC (261)	.20
	Posttraining employment	NYC (104)	.22
Job holding skills	Counselor rating	NYC (111)	.31
	Work supervisor rating	NYC (111)	.34
	Vocational skills instructor rating	OIC (260)	.15
	Remedial skills instructor rating	OIC (134)	.18
Job seeking skills	Counselor rating	NYC (111)	.22
	Work supervisor rating	NYC (111)	.31
	Posttraining employment	NYC (104)	.36
	Posttraining employment	OIC (157)	.21
Self-esteem	Counselor rating	NYC (111)	.34
	Work supervisor rating	NYC (111)	.24
	Remedial skills instructor rating	OIC (134)	.18

SOURCE: Educational Testing Service (1980).

#### CHARACTERISTICS OF THE DATA BASE

##### Completeness of Initial Coverage

According to ETS, the Standardized Assessment System was designed to provide a complete enumeration of all participants (together with appropriate controls) in all YEDPA demonstration projects. In their words (Educational Testing Service, 1980):

In a literal sense there is no "sampling" with respect to enrollees at a demonstration site since evaluation data are to be collected on the performance of all enrollees at a particular site. The control group at a particular site, however, does represent a sample from a hypothetical population that is, hopefully, similar to the enrollees with respect to important background and ability variables.

The difficult task of ensuring that data were collected in a standardized manner from all program participants was not, however, under the control of ETS. The Department of Labor had arranged for data to be collected by individual program operators; administration

TABLE A.2 ETS Estimates of Reliability and Predictive Validity of SAS Instruments

SAS Measurement	Internal Consistency (Alpha)	Predictive Validity		
		Time to Find First Job	Activity Status <sup>(a)</sup>	Level of Present Job
Vocational attitudes	.74	b	.12	.21
Job knowledge	.66	b	b	.23
Job holding skills	.56	-.16	b	.28
Work-related attitudes	.78	-.17	.10	.18
Job seeking skills	.66	-.16	b	.24
Sex stereotyping	.90	-.26	b	.16
Self-esteem	.60	-.17	b	.15

NOTE: Predictive validity estimates are for 3 months postprogram for YCD participants. Sample sizes range from 120 to 790 for validity estimates. Reliability estimates are average of values reported for participants and controls (combined n = 3,256).

<sup>a</sup>Activity status coded 0 for not working or in school, 1 for part-time work or school, and 2 for full-time work or school. It is not clear from the text how both part-time work and part-time school would be coded.

<sup>b</sup>Not significant.

SOURCE: Educational Testing Service (1980).

and execution of the data collection were not ETS's responsibility. ETS contracted to process the data supplied by the program operators (and, in a number of cases, to analyze that data).<sup>2</sup> Indeed, most ETS discussions of the SAS data base contain forceful disclaimers that "collection of all data with the Standardized Assessment System instruments remained the sole responsibility of the service delivery agents who were required to assign suitable staff at each project site for carrying out the data gathering tasks" (ETS, 1982:15, emphasis in original).

As a result of this delegation of data gathering responsibility to the program operators, there was known to be quite incomplete reporting of data. Although the precise magnitude of the incompleteness of the initial coverage was not known, ETS has informally speculated that up to 50 percent of the program participants may have been missed.

<sup>2</sup>ETS involvement in the data collection grew out of evaluation studies begun by N. Freeberg and D. Rock of Youth Career Development and Service-Mix Alternatives projects.

To investigate the characteristics of the SAS data base, we obtained a copy of the data base (minus individual identifiers).<sup>3</sup> Because data were collected from program sites on the number of new persons enrolled each month, it is possible to gain some insights into the nature and magnitude of the incompleteness of coverage. Using the "process data" provided by each site, we tabulated the total number of persons reported to be enrolled in YEDPA demonstration programs. We then tabulated data on individual participants by site to obtain an indication of the proportion of enrollees who were missing from the participant file.

As with all attempts at complete enumeration, the estimation of undercoverage is not straightforward unless there exists a valid count for the true size of the population being enumerated. In the present case, it is likely that the month-by-month counts of new program entrants were figures that program operators had readily at hand. (This results from the fact that payments to programs are tied to the number of entrants--which, of course, introduces its own potential for distortions in reporting.) If we take the reports of total enrollments at site  $i$  ( $E_i$ ) as an indicator of the total number of persons who should have been interviewed and tested, then, for any single site, the incompleteness of coverage can be represented by the ratio ( $N_i/E_i$ ) where  $N_i$  is the sample count in the participant data file for site  $i$ . For example, if a site said it enrolled 2,000 youth but only 1,200 respondents from that site could be located in the participant file, then the coverage rate could be said to be 1,200/2,000 or 60 percent. (Note, however, how inflating of enrollment figures by program operators or mechanical errors in data entry or matching might bias this estimate.)

If all program sites accurately reported enrollment data, we might then make a confident estimate of the completeness of coverage by summing across sites ( $\sum N_i / \sum E_i$ ). The data do not, however, comply so readily with our wishes. Incompleteness affects not only participant data, but also the process data.

Analysis of the process data collected from individual program sites (shown in Table A.3) reveals that the majority of program sites did not provide data on their program operations. This can be detected within the ETS data base because site codes appear on both respondent records and site records. We thus know (assuming the site codes have

---

<sup>3</sup>Three observations should be made about technical aspects of the data sets. First, the documentation provided with the data sets was not always adequate. Second, although the data have a hierarchical structure (there are respondents within sites within programs), the data sets are not designed to encourage analyses that make use of that hierarchical structure. Third, no procedural history exists for the data gathering. Thus it is unclear how many sites were contacted for data, how many sites provided data that was judged "suspicious," how ETS "winnowed" the data set to eliminate "suspicious" data, and so forth.

TABLE A.3 Cross-tabulation of Availability of Site "Process" Data in YEDPA/SAS Data Base by Availability of Participant Data

Participant Data	Site "Process" Data	
	Missing	Reported
No data	Unknown (n = 0) <sup>a</sup>	24 sites (n = 0) <sup>a</sup>
Data on one or more participants from site	292 sites (n = 21,839 part.) (n = 8,774 cont.)	166 sites (n = 12,733 part.) (n = 6,836 cont.)
Total reported enrollment	Unknown	29,272 enrollees

NOTE: Numbers in parentheses are total number of participant (part.) and control (cont.) cases in respondent data base for those sites.

<sup>a</sup>Sites provided no participant or control data.

been accurately recorded) that 458 program sites provided some participant data to ETS. Of those 458 sites only 166 provided "process" data on site operations. Not only do these "missing" sites constitute the majority of identifiable sites, but they also account for the majority of the respondents whose data were supplied to ETS (30,613 of 50,182 respondents in the data base came from sites that did not report "process" data).

Some partial information on coverage can still be gleaned from Table A.3. Note, for example, that the sites providing process data reported a total enrollment of 29,272 participants. For these same sites we find only 12,733 cases in the participant file (plus 6,836 control cases). For these sites, the coverage estimate would be 12,733/29,272 or 43 percent. (It is, of course, a leap of faith to assume that this same percentage would apply to the sites that did not provide process data, but it is not inconsistent with the informal "guesstimates" made by ETS personnel who were familiar with the SAS data collection.)

#### Sample Attrition

Subsequent to the initial data gathering, a series of data collection steps were planned for each participant (and control) in the SAS data base. Interviews were to be conducted at program completion, the Job Knowledge and Attitudes battery was to be re-administered, subsequent survey interviews were to be conducted 3 months and 8 months after program completion, and data were to be gathered from employers, counselors, and so on.

In all longitudinal surveys, one expects some reduction in the numbers of respondents from whom data can be collected in each succeeding wave of data gathering. Such sample loss may compromise the representativeness of the remaining sample (save in the rare case when sample loss is effectively random). Such attrition, however, is a fact of life that social researchers have to live with. Respondents move and are untraceable, they lose their patience with the researchers' persistent inquiries, and so forth. Nonetheless, it is not unknown for well-conducted surveys to obtain responses from 80 percent (or more) of the original interviewees after a period of months (and even years). Systematic follow-up and dogged determination to find the "movers" and persuade the "refusers" have resulted in some remarkably low attrition rates in long-term follow-up studies, even with very youthful populations (see Sewell and Hauser, 1975, and descriptions of the Continuous Longitudinal Manpower Survey and National Longitudinal Survey in Appendix D).

Table A.4 shows the various stages of the SAS data collection and the attrition that occurred over time. It will be seen from Table A.4 that the attrition in the SAS data collection base is sufficiently high to engender skepticism about claims that the follow-up data provide reasonable estimates of the postprogram experiences of the youth from whom data were initially collected (not to mention the entire population of participants).<sup>4</sup>

Looking at the data base in its entirety (i.e., including all respondents from whom any data were collected) reveals that at 8 months after program completion, there is no interview data on the majority of program participants and controls (see Table A.4). Even at 3 months postprogram, the attrition losses amount to almost half of the original sample (45 percent of program participants and 49 percent of controls who provided initial data did not provide interview data at 3 months postprogram).

Sample attrition is clearly at a level at which serious doubts must arise about whether the results obtained from the follow-up samples can be generalized. Even for postprogram measurements, attrition rates are rather high: 32 percent of participants lack postprogram survey data.<sup>5</sup>

---

<sup>4</sup>In fact, sample attrition in the data base is higher than that reported in the ETS analyses. This difference in numbers arises because ETS eliminated approximately 11,000 cases from their published figures as a result of their "winnowing" of the sample to exclude "suspicious" data.

<sup>5</sup>It is possible that some portion of this "loss" may be attributable to youths who dropped out of the program (rather than "completers" for whom data are missing). If this is so, the situation may not be quite so bleak as it seems, since one could attempt to cast an analysis in terms of the effects of program completion (rather than mere enrollment in a YEDPA program).

TABLE A.4 SAS Data Collection Program: Instruments, Schedule, and Samples Obtained

Instrument	Administration Schedule	Population from Whom Data Collected	N	
			Participants	Controls
Individual Participant Profile section 1 (demographic and SES characteristics) <sup>a</sup>	Beginning of program	Participants and controls	34,572	15,610
STEP (reading test)	Beginning of program	Participants and controls	30,191	12,998
Pretest (7 knowledge and attitudinal measures) <sup>a</sup>	Beginning of program	Participants and controls	31,732	13,610
IPP section 2 (program termination data)	End of program	Participants	not available	--
Posttest (same as pretest) <sup>a</sup>	End of program	Participants and controls	24,580	9,314
Program completion survey (interview)	End of program	Participants	23,354	--
Counselor's rating	End of program	Program counselors of participants	22,232	--
Work supervisor's rating	End of program	Work supervisors of participants	11,068 <sup>b</sup>	--
Control-Group Status Survey (interview)	End of program	Controls	--	6,471
Three-month follow-up survey (interview) <sup>a</sup>	Three months after end of program	Participants and controls	18,870	7,948
Three-month employer rating	Three months after end of program	Employers of participants and controls	2,341 <sup>b</sup>	523 <sup>b</sup>
Eight-month follow-up survey (interview) <sup>a</sup>	Eight months after end of program	Participants and controls	14,906	5,461
Eight-month employer rating	Eight months after end of program	Employers of participants and controls	2,023 <sup>b</sup>	506 <sup>b</sup>
Project and Process Information Questionnaire (data on program environment and processes)	Once during program operation	Program personnel	12,733 <sup>c</sup>	6,836 <sup>c</sup>

<sup>a</sup>Instruments for which N's are included in Figure A-1.

<sup>b</sup>Instrument not applicable to all respondents.

<sup>c</sup>N's shown are number of youths in sites who returned Process Information Questionnaire.

The magnitude of the attrition in the ETS data base makes one wonder why it was not anticipated either by the agency, the contractor, or the reviewers at the Office of Management and Budget (OMB) who were responsible for approving government data-gathering activities. We would note, in this regard, that in 1977 OMB announced an explicit standard for response rates, which (temporarily) replaced its non-numerical standard of adequacy (Office of Management and Budget, 1977):

It is expected that data collections for statistical purposes will have a response rate of 75 percent. Proposed data collections having an expected response rate of less than 75 percent require a special justification. Statistical data collection activities having a response rate below 50 percent should be terminated. Proposed statistical data collections having an expected response rate of less than 50 percent will be disapproved.

Clearly, under such a standard, this data collection would not have been approved (or re-approved) if the attrition rates for its follow-up measurements had been accurately anticipated.

To illustrate the cumulative impact of the incompleteness of coverage discussed previously and the sample attrition that occurred over time in the SAS data base, Figure A.2 graphs an estimate of the size of the target sample of participants for the SAS data collection (using the 43 percent coverage rate computed for sites reporting process data) together with the sample sizes obtained for both participants and controls over the course of the data gathering. The effects of undercoverage and sample attrition are quite dramatic.

Because this sample attrition was so great we undertook some exploratory analyses to assess its effect on the composition of the SAS samples over time. Table A.5 presents tabulations of a variety of economic and social characteristics for respondents who provided interview data at entry and at 3 and 8 months postprogram. A few modest trends are evident from Table A.5. The proportion of high school dropouts declines from 25 to 20 percent, while the proportion of persons receiving public assistance shows a modest rise (42 to 47 percent for participants and 38 to 42 percent for controls). While there are notable changes and a very large number of "significant" differences (given the large sample sizes) between those who continued to provide data and those who were lost to attrition, it was surprising how modest the changes were. When a parallel analysis was performed on the Job Knowledge and Attitude measurements (see Table A.6), the results were equally unprovocative.

#### Reliability and Validity of Instruments

Job knowledge and attitude measures figure prominently in many of the evaluation studies conducted under YEDPA. The rationale behind the use of such instruments is that they measure traits that (1) the programs change and that (2) are important in helping youths find



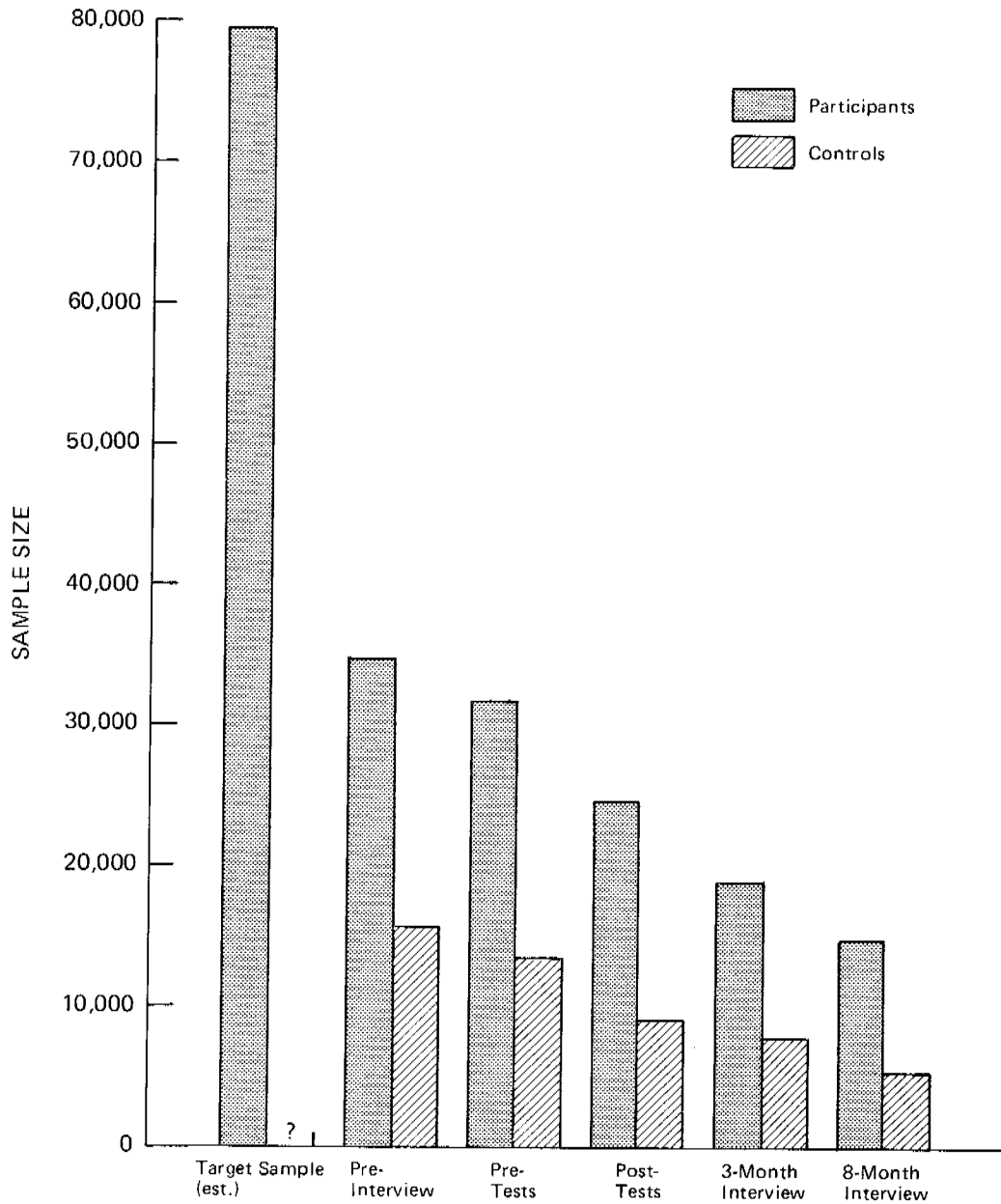


FIGURE A.2 Sample coverage and attrition (Standardized Assessment System).

employment. Since these are relatively inexpensive data to collect, there is some reason to favor such a strategy--particularly if one suspects that the effects of training on employment may be unusually subtle or delayed in arriving. This strategy, of course, depends on the measures being adequate in the sense of being replicable so that repeated measurements are relatively stable and in their being

TABLE A.5 Social and Demographic Characteristics of SAS Samples

Respondent Characteristic	Sample	Stage		
		Entry %	3-Month Follow-up %	8-Month Follow-up %
Female	Participant	51.9	54.3	54.6
	Control	51.1	52.5	53.2
High school dropout	Participant	25.5	21.6	20.5
	Control	25.4	24.5	20.4
Income 70% of standard	Participant	66.3	61.7	63.7
	Control	56.7	56.4	59.8
Welfare recipient	Participant	42.7	45.2	47.6
	Control	38.6	41.1	42.1
Race/ethnicity				
Black	Participant	56.2	58.5	58.1
	Control	53.4	55.2	54.5
Hispanic	Participant	21.9	22.4	23.1
	Control	26.7	29.2	29.1
White	Participant	19.5	16.8	17.0
	Control	17.7	14.3	15.2
Limited english	Participant	7.6	8.1	7.6
	Control	10.2	9.4	8.9
Has children	Participant	11.3	10.6	11.7
	Control	8.5	7.9	8.4
Criminal offender	Participant	8.2	6.7	7.1
	Control	11.5	10.0	12.0
Previous CETA participant	Participant	30.0	30.8	30.7
	Control	25.5	25.7	25.9

NOTE: Less than 1 percent of data records were inconsistent, e.g., the respondent was a "control" but the 3-month follow-up flag indicated the respondent had completed a "participant" follow-up survey. These records were excluded from this analysis.

SOURCE: Derived by tabulating data for every fifth record in ETS data base, i.e., 20 percent subsample of data base.

reasonable proxies for the more difficult to observe outcomes. The former condition is generally referred to under the rubric of reliability, the latter as validity (of one sort or another).

Since the Standardized Assessment System was launched with some expressed trepidations about the suitability of such measures to the YEDPA population, it is important to seek evidence within the data base as to whether these conditions are met by the SAS measurements. SAS provides the opportunity for making (test-retest) reliability estimates

TABLE A.6 Job Knowledge and Attitudes and Other Pretest Scores (at entry) of Respondents Giving Interviews at Entry and 3 and 8 Months Postprogram

SAS Measurement	Sample	Stage			Standard Deviation of Scale
		Entry	3-Month Follow-up	8-Month Follow-up	
Vocational attitudes	Participant	20.5	20.5	20.4	4.5
	Control	20.2	20.2	20.6	
Job knowledge	Participant	21.6	21.8	21.6	4.2
	Control	21.4	21.3	21.7	
Job holding skills	Participant	30.4	30.6	30.5	2.7
	Control	30.2	30.2	30.3	
Work-related attitudes	Participant	40.8	48.2	48.1	6.8
	Control	47.9	47.8	48.5	
Job search skills	Participant	11.7	11.8	11.7	3.2
	Control	11.5	11.3	11.7	
Sex stereotyping	Participant	45.4	45.1	45.3	8.2
	Control	45.0	44.6	45.0	
Self-esteem	Participant	36.3	36.4	36.3	3.2
	Control	35.9	35.9	36.2	
Reading ability (STEP)	Participant	15.0	15.0	15.1	4.6
	Control	14.5	14.6	14.9	

NOTE: Standard deviation of scale is computed from data for all controls and participants.

SOURCE: Derived by tabulating data for every fifth record in ETS data base, i.e., 20 percent subsample.

for these scales, since the same battery was administered preprogram and postprogram to the untreated controls. Although one can expect true temporal change to affect the cross-temporal correlations between two measures of a trait such as self-esteem or work-related attitudes, one would expect a certain amount of stability in these traits. After all, if people varied widely from day to day on these traits it is not (easily) conceivable that the measure would be helpful in predicting relatively stable social behaviors, such as employment or other vocational behaviors.

A series of analyses reported in Tables A.7 through A.10 examine some of the properties of these scales. In Table A.7, the zero-order correlation of each scale measured preprogram and postprogram is

TABLE A.7 Test-Retest Reliability for SAS Measurements (computed from 20 percent sample of ETS data base)

SAS Measurement	Zero-Order Correlation Over Time <sup>a</sup>	
	Controls	Participants
Vocational attitudes	.604	.602
Job knowledge	.527	.505
Job holding skills	.460	.386
Work-related attitudes	.604	.610
Job search skills	.572	.538
Sex stereotyping	.631	.643
Self-esteem	.462	.388
(N) <sup>b</sup>	(1,644)	(4,443)

NOTES: Test-retest reliability will be affected by "true change" in respondents. Since participants are enrolled in programs designed to change their attitudes and knowledge, reliability estimates for this group should be treated with caution.

<sup>a</sup>Measurements made using identical instruments preprogram and postprogram.

<sup>b</sup>Minimum sizes of samples from which estimate was made.

TABLE A.8 Correlations Between Reading Scores and SAS Measurements of Job Attitudes and Knowledge (computed from 20 percent sample of ETS data base)

SAS Measurement <sup>a</sup>	Correlation with STEP Reading Score	
	Participant	Control
Vocational attitudes	.445	.509
Job knowledge	.447	.467
Job holding skill	.288	.354
Work-related attitudes	.446	.383
Job search skill	.569	.578
Sex stereotyping	.279	.241
(N) <sup>b</sup>	(5,603)	(2,258)

<sup>a</sup>All measurements made during pretest.

<sup>b</sup>Maximum sizes of samples upon which any reported correlation is based.

TABLE A.9 Zero-Order and Partial Correlations (with reading) between SAS Pretest Measurements (calculated from 20 percent sample of SAS data base)

SAS Measurement	Type of Correlation	SAS Measurement							SE
		VA	JK	JHS	WRAI	JSS	SS		
Vocational attitudes (VA)	Zero-order	--	.427	.333	.564	.472	.312	.278	
	Partial	--	.286	.242	.455	.297	.219	.184	
Job knowledge (JK)	Zero-order	.405	--	.462	.440	.582	.224	.303	
	Partial	.261	--	.394	.302	.449	.118	.213	
Job holding skill (JHS)	Zero-order	.347	.431	--	.448	.443	.170	.402	
	Partial	.240	.343	--	.376	.360	.101	.354	
Work-related attitudes (WRAI)	Zero-order	.528	.419	.410	--	.496	.331	.442	
	Partial	.416	.289	.317	--	.330	.241	.373	
Job search skill (JSS)	Zero-order	.474	.545	.447	.490	--	.250	.368	
	Partial	.299	.406	.339	.330	--	.119	.271	
Sex stereotyping (SS)	Zero-order	.275	.198	.152	.287	.242	--	.138	
	Partial	.189	.106	.078	.206	.127	--	.068	
Self-esteem (SE)	Zero-order	.299	.319	.421	.446	.411	.152	--	
	Partial	.179	.209	.353	.357	.287	.077	--	

NOTE: All measurements are from 20 percent sample of ETS pretest data. Partial correlations control for STEP reading scale score. Step-wise case deletion was done (N = 2,196 controls and 5,464 participants). Correlations for program participants are shown above diagonal; those for controls are below.

TABLE A.10 Estimates of Predictive Validity of SAS Attitude and Knowledge Measurements (computed from 20 percent sample of ETS data base)

SAS Measurement at Program Completion	Sex	Correlation with Activity Status <sup>a</sup>	
		3 Months Postprogram	8 Months Postprogram
Vocational attitudes	Male	(.048)	(.056)
	Female	.078	(.042)
Job knowledge	Male	.085	(.042)
	Female	.057	(.042)
Job holding skills	Male	.111	.075
	Female	(.043)	(.031)
Job seeking skills	Male	.097	(.042)
	Female	.060	.055
Work-related attitudes	Male	.068	(.057)
	Female	.107	.082
Self-esteem	Male	.060	(-.003)
	Female	.092	(.043)
Sex stereotyping	Male	-.054	(-.029)
	Female	.050	(.013)
N	Male	1,080	714
	Female	1,326	935

NOTE: Estimated coefficients in parentheses are not reliably different from 0 (at  $p < .05$ ). Correlations are derived from 20 percent subsample of YEDPA participants in ETS data base. Respondents were included only if they were coded as a participant in IPP profile and if the data flag for the 3-month follow-up indicated they had completed a participant follow-up survey (not a control survey). In a small number of cases (281 of 50,182), those two indicators are inconsistent; those cases were excluded from this analysis.

<sup>a</sup>Activity status is coded 1 if respondent is in a full-time job or is a full-time student; status is coded 0 otherwise.

reported for program participants and controls.<sup>6</sup> All of the test-retest correlations are within the range of approximately 0.4 to 0.6. While these correlations are not insubstantial, neither would

<sup>6</sup>Obviously, the reliability estimates for the control groups are most relevant since the controls did not participate in YEDPA's programs that were designed to change participant's attitudes, knowledge, behavior, and so forth. However, as Table A.7 shows, reliability estimates for program participants are quite similar to those for controls.

they be thought to indicate an extremely robust measurement. Indeed, if one were to assume that measurement errors (both random and systematic) did not contaminate these data, these estimates would suggest a great deal of variation over time in young people's knowledge of and attitudes toward jobs, their self-esteem, and the extent to which they sex stereotype the occupational world. This could, of course, be the case. But it is also plausible that a relatively large component of measurement error may be distorting the measurements.

Overall, the self-esteem scale and the job-holding skill scale show relatively low cross-temporal correlations, while the sex stereotyping and vocational attitudes scales show correlations of 0.6 or better. (In the case of the sex stereotyping scale, one suspects that this relatively high estimate of reliability may derive, in part, from the fact that all items were presented in the same format and scored in the same direction.) These scales also show a high correlation with reading ability. Table A.8 presents the correlations between each of these scales (measured at pretest) and the STEP reading scale scores. These correlations range from a low of .241 for sex stereotyping to a high of .578 for the job search skill scale.

While one might be tempted to dismiss some of these high correlations with reading ability as "artifacts," for some purposes the correlation is as one would want. The ability to read a job advertisement is an essential component of "job search skills." It is not, however, the case that such a simple argument can be made to defend these correlations in every instance. There is no prima facie case to be made for a correlation between the attitude measures and reading--although there are more than enough plausible paths for indirect causation to account for this correlation. It is important to keep in mind that reading (and a myriad of other unmeasured traits) may play a role in accounting for the zero-order test-retest reliabilities. Potential correlated measurement errors also bedevil all attempts to understand test-retest reliabilities.

Some evidence of the construct validity of the various SAS measurements may be gleaned from Table A.9. As intended by the SAS designers, all of the measures are positively intercorrelated. This is true even when a simplistic attempt is made to account for confounding effects of reading ability on all of the scales. The strongest correlations found for the SAS measures are between scales that measure

---

It should be realized that test-retest correlations such as those shown in Table A.7 are affected by both true change in the respondents and by measurement errors. If one wishes to use measures like the SAS assessment battery as proxies for (unmeasurable) long-term outcomes (e.g., lifetime earning potential and employability), however, instability, per se, may be an important consideration. If a characteristic like work-related attitudes, for example, naturally varies to such a degree that test-retest correlations approach zero over a short period of time (in the absence of measurement error), then even a perfectly reliable measurement of this characteristic would be of doubtful utility in most program evaluations.

similar or related traits, e.g., vocational attitudes and work-related attitudes, or job search skills and job knowledge. Conversely, correlations between the sex role stereotyping measures and job knowledge factors are low.

### Predictive Validity

Given the aim of the YEDPA programs, a key validity test for any scale would be its ability to predict which YEDPA youth would stay in school or find full-time employment and which would not. Several skirmishes have been made with this analysis and Table A.10 reports the simplest of them. (Its outcome, however, is little different from the more complicated analyses.)

For all program participants (in 20 percent sample) who provided the requisite data at 3 months ( $n = 2,406$ ) and at 8 months ( $n = 1,649$ ) postprogram, a score of 1 was assigned if (at follow-up) the respondent reported being either in school full time or working full time. A score of 0 was assigned otherwise. In the crudest analysis (reported in Table A.10) the 0-order correlation between this dichotomous "activity variable" and each of the scales from the SAS battery was calculated.<sup>7</sup> This was done separately for males and females to allow the effects of potential differences in child-care responsibilities to appear.

It will be seen from Table A.10, that there were some "significant" correlations between job knowledge and attitude scores and whether a youth was "occupied" or "unoccupied," however, the magnitude of these correlations was not substantial. The correlations for the SAS data base are considerably below those found for the NYC and OIC samples reported by ETS in their 1980 report on SAS (Educational Testing Service, 1980). They are even lower than the correlations (.10) reported by ETS from the Youth Career Development sample.

The extremely low predictive validity of the SAS measures raises questions about the meaningfulness of program evaluations that rest their verdicts of program effectiveness on such measurements. As Chapters 4 through 8 have shown, such studies are not uncommon in the YEDPA literature.

### Inter-site Variations

The shortcomings of the aggregate SAS data base invite the question: Is the data base uniformly riddled with such problems? It

---

<sup>7</sup>This analysis is somewhat crude, but it illustrates the point in a straightforward manner (and it is analogous to analyses reported in ETS, 1980). It should be noted, however, that because the criterion variable is dichotomous, the obtained correlations will understate somewhat the extent of the relationship.



TABLE A.11 Follow-up Rates for 10 Randomly Selected Sites in SAS Data Base

Postprogram Follow-up Stage <sup>a</sup>	Sample	Follow-up Rates				No Sample <sup>b</sup>
		0-24%	25-49%	50-74%	75+%	
3 months	Participants	1	3	2	4	--
	Controls	1	3	1	2	3
8 months	Participants	5	2	2	1	--
	Controls	4	2	1	0	3

NOTE: Ten sites were selected using a random number table from among all sites in the ETS data base having an "n" of at least 25 (controls + participants) at Wave 1. N's for samples whose rates are shown above range from 24 to 167.

<sup>a</sup>Follow-up rate is a percentage of all respondents at site for whom there is any 3-month (or 8-month) interview data (as indicated by "flags" set in the data base to indicate presence or absence of these data).

<sup>b</sup>Three sites had no control groups.

is possible in theory, of course, for an aggregate outcome such as the one reported here to be composed of some very fine data gathering operations and some very poor ones. While the aggregate result would not be impressive, it still might be possible to isolate a sizable subset of the data base upon which a convincing analysis could be performed.

To assay this possibility, we selected 10 sites at random from the SAS data base and ascertained the distribution of attrition rates across sites. We restricted the universe of potential sites for this analysis to sites that had a minimum of 25 respondents (controls and participants) at the initial data collection. For each of these sites, we then computed the follow-up rates at 3 and 8 months postprogram. The distribution of follow-up rates across the 10 sites is shown in Table A.11. It will be seen that at 3 months postprogram four sites had follow-up rates for participants of 75 percent or higher. For the control samples, only two sites had such high follow-up rates.

While the attrition analysis at 3 months is somewhat encouraging, the results at 8 months are quite disappointing. Only one site maintained a 75 percent follow-up rate for participants at 8 months, and no site attained this rate for its control samples.

In addition to the analysis of attrition rates, we also attempted to assay the distribution across sites of the predictive validity of the SAS attitude and knowledge measurements. Here again, we selected 10 sites at random from the SAS data base. This time, however, we restricted our analysis to sites that had a minimum of 100 program participants from whom data had been obtained at 3 months postprogram. This was done to provide an adequate sample size for calculating the correlation coefficients between the (immediate) postprogram SAS measurements and the participants' "activity status" at 3 months

TABLE A.12 Distribution of Predictive Validities of SAS Attitude and Knowledge Measurements For 10 Randomly Selected Sites

SAS Measurement	Magnitude of Correlation with Activity Status at 3 Months Postprogram <sup>a</sup>						Not Computable <sup>b</sup>
	-.20 to -.11	-.10 to -.01	0.00 to 0.09	.10 to .19	.20 to .29	.30+	
Vocational attitudes	0	1	5	2	1	0	1
Job knowledge	1	0	6	2	0	0	1
Job holding skills	0	0	5	4	0	0	1
Job seeking skills	1	1	2	4	1	0	1
Work-related attitudes	0	1	3	2	3	0	1
Self-esteem	0	1	3	4	1	0	1
Sex stereotyping	0	2	4	2	1	0	1
Summary Frequences	2	6	28	20	7	0	

NOTES: Ten sites were selected using a random number table from among the 66 sites in the ETS data base that had some 3-month interview data on at least 100 participants. N's for correlations within each site range from 70 to 186.

Respondents were included in this analysis only if they were coded as a program participant in IPP profile and if the data flag for the 3-month follow-up indicated they had completed a participant follow-up survey (not a control survey). In a small number of cases in the overall data base (281 of 50,182), these two indicators are inconsistent; these cases were excluded from this analysis. Attitude and knowledge scores are from postprogram measurement wave (i.e., "zero" months postprogram).

<sup>a</sup>As in Table A.10 the criterion being predicted is "activity status," which is coded 1 if respondent is in a full-time job or is a full-time student (it is coded 0 otherwise). Students enrolled in school at or below the tenth grade were excluded from the analysis since such students frequently have little discretion about their status as a result of compulsory education laws.

<sup>b</sup>One of the 10 randomly selected sites did not administer the SAS postprogram battery, and hence, correlations could not be computed for this site.

postprogram. Table A.12 presents the results of this analysis. (See the notes to Table A.12 for definitions of sample selection criteria and the activity status variable.)

It will be seen from Table A.12 that no predictive validity for any measurement at any site exceeded 0.30. The vast majority of correlations (48 of the 60 that could be calculated) were in the range 0.0 to 0.20. Indeed, over half of the coefficients we calculated (36 of 60) were less than 0.10.

While it would be a mistake to overgeneralize based on data from such a small number of sites, these data on attrition and measurement validity do not encourage the belief that there exist a sizable number of sites in the SAS data base that gathered high-quality data (where quality is indicated by the attrition of the sample and the predictive validity of the measurements).

#### REFERENCES

##### Educational Testing Service

1980 The Standardized Assessment System for Youth Demonstration Projects. Youth Knowledge Development Report No. 1.6. Washington, D.C.: U.S. Department of Labor.

1982 Demonstration Programs for Youth Employment and Training: The Evaluation of Various Categories of YEDPA Program Sites. Princeton, N.J.: Educational Testing Service.

##### Office of Management and Budget

1977 Memorandum to heads of executive departments, February 17, 1977.

##### Sewell, W., and R. Hauser

1975 Education, Occupation and Earnings. New York: Academic Press.

##### Taggart, R.

1980 Youth Knowledge Development: The Process and the Product. Unpublished manuscript.