DOCUMENT RESUME

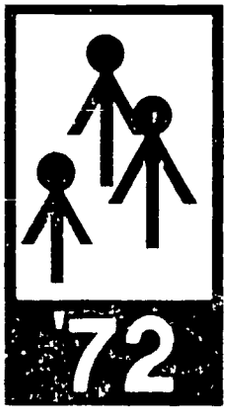
ED 151 398 TM 006 859

AUTHOR	Peng, Samuel S.; And Others
TITLE	National Longitudinal Study of the High School Class of 1972. Differences Between High School Graduates and Dropouts in Participation Patterns and Response Quality in Survey Studies.
INSTITUTION	Research Triangle Inst., Durham, N.C. Center for Educational Research and Evaluation.
SPONS AGENCY	National Center for Education Statistics (DHEW), Washington, D.C.
PUB DATE	Feb 77
CONTRACT	OEC-0-73-6666
NOTE	23p.
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	*Attrition (Research Studies); *Comparative Analysis; Data Collection; *Dropouts; Followup Studies; Graduate Surveys; *High School Graduates; *Longitudinal Studies; *National Surveys; Participation; Question Answer Interviews; Questionnaires; Research Methodology; Research Problems; Senior High Schools; Simulation; Statistical Data; Student Characteristics
IDENTIFIERS	Mail Surveys; *National Longitudinal Study High School Class 1972

ABSTRACT

A group of high school self-reported dropouts and a group of simulated high school dropouts were compared with graduates in their participation rates, efforts required for data collection, and the overall quality of their responses in the National Longitudinal Study of the High School Class of 1972. The results indicate that dropouts are less likely to return mail questionnaires, have lower retention rates in a longitudinal study, need more effort to trace their residence and to complete editing checks, and generally provide data of lower quality. The results also indicate that no substantial differences in data quality exist between those dropouts and graduates who were interviewed. It is thus suggested that the interview approach is preferable as far as data quality is concerned. The combined use of mail survey, extensive telephoning, and interviews would also be advisable. In either case, questionnaires should be simple and straightforward. A complicated and lengthy questionnaire would probably discourage dropouts from completing the questionnaire. (Author/EVH)

NATIONAL LONGITUDINAL STUDY of the High School Class of 1972



DIFFERENCES BETWEEN HIGH SCHOOL GRADUATES AND DROPOUTS IN PARTICIPATION PATTERNS AND RESPONSE QUALITY IN SURVEY STUDIES

Samuel S. Peng
J. P. Bailey, Jr.
G. H. Dunteman

Prepared for the
National Center for Education Statistics
Office of the Assistant Secretary for Education
Department of Health, Education, and Welfare
Contract No. OEC 0 73-6666

by
Center for Educational Research and Evaluation
Research Triangle Institute
Research Triangle Park, North Carolina 27709

February 1977

Abstract

A group of high school self-reported dropouts and a group of simulated high school dropouts were compared with graduates in their participation rates, efforts required for data collection, and the overall quality of their responses in a longitudinal survey. Results indicated that dropouts were less likely to return mail questionnaires; had lower retention rates in a longitudinal study; needed more effort to trace their residences and to complete editing checks; and generally provided data of lower quality. Results also indicated that no substantial differences in data quality existed between those dropouts and graduates who were interviewed. It is thus suggested that the interview approach is preferable as far as data quality is concerned. A combination of mail survey, extensive editing phone calls, and interviews would also be advisable. In either case, questionnaires should be simple and straightforward. A complicated and lengthy questionnaire would probably discourage dropouts from completing the questionnaire.

Foreword

The National Longitudinal Study of the High School Class of 1972, a survey initiated by and conducted for the National Center for Education Statistics, began in the spring of 1972 with over 1,000 in-school group administrations of survey forms to a sample of approximately 18,000 seniors. In the follow-up surveys, the sample was augmented by almost 5,000 additional students from sample schools that were unable to participate in the base-year survey.

The data collected from the in-school and two follow-up surveys have been merged and processed. Results are being presented in a series of reports, designed to highlight selected findings in educational, career, and occupational development. This report contains information about the differences between high school graduates and dropouts in participation rates, data collection efforts, and response quality.

Continuing follow-up requests for data from these individuals are planned through 1979 and perhaps beyond. This series of repeated observations will permit the examination of the relationships between schooling, work, and other experiences to subsequent career choices as well as educational and labor force participation of each of the selected individuals. Such information and the resultant analyses are important to those engaged in formulating legislative proposals and educational policy.

Francis Corrigan, Director
Division of Multi-Level
Education Statistics

Elmer F. Collins, Chief
Longitudinal Studies Branch

Table of Contents

List of Tables

List of Figures

Differences Between High School Graduates and Dropouts in Participation Patterns and Response Quality in Survey Studies

I. BACKGROUND OF THE STUDY

This study was designed to seek answers to questions concerning the data collection methods intended for use in future surveys of the National Longitudinal Study (NLS). Although the resultant information can have broader generalizations beyond NLS, a brief description of the study background should be helpful in giving a better understanding of the problem

. The National Longitudinal Study is a continuing project monitored and primarily funded by the National Center for Education Statistics (NCES). This project represents a number of agencies and offices within the Department of Health, Education, and Welfare with interests in the long-range effects of educational policy. The focus of the NLS is on the educational, vocational, and personal development of high school students, and the personal, familial, social, institutional, and cultural factors that contribute directly or indirectly to that development. The general purpose is to establish a factual basis for verifying and refining federal policy concerned with maximizing individual access to educational and vocational opportunity, with improvement of the general educational system as it impinges on young people, and with positive impact toward aiding young people to assume a productive, satisfying, and wholesome adult role in society. A secondary purpose is to extend the general scientific knowledge of human development in the important years covering the transition from high school to adult careers.

The NLS began with a stratified, two-stage probability sample of over 21,000 students representing the graduating high school class of 1972, with follow-up to date in 1973, 1974, and 1976. For various reasons the 1972 study was restricted to those individuals who were high school seniors in the spring of 1972. The study design did not include high school dropouts, an especially important group if policy-oriented research is to deal, as it should, with the broad dynamics and wide variations of equal opportunity by race, class, and sex.

As another projected part of the National Longitudinal Study, substantial replication of the 1972 study is planned with a later cohort, probably the High School Class of 1980, with the important addition of including individuals who had been members of that class but who left school before graduating. These dropouts will be included so that their post-school development may be compared with that of those who do graduate, and to broaden the population base to which generalizations concerning educational effects can be applied.

The inclusion of a dropout component in the study, however, requires additional considerations for data collection methods because of the following questions: Are dropouts more difficult to locate, less likely to complete certain types of instruments successfully, and less cooperative than nondropouts? Previous research has not provided adequate empirical bases for answering these questions. References can be drawn only from some related studies. For example, it has been found that less educated people have higher nonresponse rates in mail survey (Moser & Kalton, 1972) but are the most likely to be located (Crider & Willits, 1973), and that less educated people are more likely to give contradictory responses (Lenski & Leggett, 1960). It has also been found that poor achievers in college have a higher nonresponse rate (Neuss, 1943), and that more nonrespondents are from families of lower socioeconomic background (Vincent, 1964). None of these studies, however, have systematically investigated the response patterns of high school dropouts in comparison with graduates. It is within this context that this study was designed.

II. GENERAL APPROACH FOR THE STUDY

To answer the above questions, one could identify high school dropouts and graduates through school records, and then actually conduct a study to compare, for example, their participation rates and response quality. This approach would be preferable to the approach used in this study. It would, however, be very costly.

The approach used in this study was to use available data from the 1972 NLS study. In this study, there were 127 sample members who reported that they left high school without completion by the fall of 1973. This group of students, labeled as self-reported dropouts, constitutes a useful study group except for the following constraints: (1) the n (sample size) may be too small for national generalizations; and (2) they may not be typical high school dropouts considering their leaving school only a few months before graduation. They probably have higher achievement and different school attitudes than those who left school before tenth grade.

To compensate for the shortcomings of the self-reported dropouts, a group of dropouts was simulated from the NLS sample. It was felt that a well-simulated group would yield as reliable estimates as an actual group and at a low cost.

This study thus includes three groups for comparisons: (1) self-reported dropouts; (2) simulated dropouts; and (3) graduates. The first group was used in a sense to validate the second group; that is, their characteristics and survey response qualities provide a basis for evaluating how close the simulated dropouts are to the actual dropouts. The simulation procedures are described in the next section.

One consideration in the conduct of this study was that the information about tracing effort and editing requirements was not in the data file. In addition, the data comprising the current NLS data files may not truly represent the quality of the raw questionnaire data because student responses to some key questions in each survey were manually edited and checked by phone calls to respondents before being keyed into data files. Consequently, the observed differences between dropout and graduate groups regarding data quality, using the data files only, could be underestimates of actual differences. Hence, the use of the original or unedited data is desirable for comparisons on data quality. However, checking the original responses meant going back to microfilm records of original questionnaires which, if the entire NLS sample were involved, would be extremely expensive. Therefore, subsamples of 100 cases each were randomly drawn from each of the dropout and graduate groups to obtain supplemental information for comparisons on tracing effort, editing requirements, and data quality.

III. SIMULATION OF HIGH SCHOOL DROPOUTS

The question addressed in this section is: Which individuals should be classified as simulated dropouts? Before answering this question, a description of the general profile of high school dropouts—those students leaving school without completion—would be helpful.

Recent estimates of the number of high school dropouts range generally from about 20 to 30 percent of the school-age population, depending upon the group to whom the label is applied and the population base used. Variations by state and region are, of course, large. In any case, all studies support Dentler and Warshauer's (1965) description of the model dropout as "a low school achiever, usually below grade level for his age. He is a member of a low-income family in which parents have low educational attainment." A national study conducted by the Survey Research Center (SRC) at the University of Michigan for the U.S. Office of Education found that whereas about 40 percent of the tenth grade students with IQs below 90 dropped out of school, fewer than 10 percent of those with scores over 125 left school

(Bachman, Green, & Wirtanen, 1971). In the same study, it was found that 40 percent of the students who were held back a grade level, versus 10 percent of those who remained in grade, subsequently dropped out of school. The relationship between academic grades and dropping out was even stronger. The SRC study also shows that whereas about 30 percent of the tenth grade students in the lowest socioeconomic quartile dropped out of high school, fewer than 10 percent from the highest quartile dropped out.

Dropouts tend to come from rural areas and large cities. The dropout rate in rural areas is almost twice as large as that in the urban fringe (U.S. Bureau of Census, 1973). Dropouts also tend to have a stronger negative attitude toward school (e.g., school is very boring) and to place less value on academic achievement. (Bachman, Green, & Wirtanen, 1971). Dropouts also tend to be more rebellious in school than nondropouts. The SRC study estimated that nearly one-half of those students who "often" engaged in rebellious behavior in school later became dropouts.

In summary, dropouts are a group of students who are generally poor, low achievers, and behind in grade level. They have high alienation toward school and place little value on academic achievement.

In light of what is known about the characteristics of high school dropouts, the next step was to identify simulated dropouts from the 1972 study sample. The 127 self-reported dropouts were separated first from the sample members, and then simulated dropouts were identified from the remaining sample members. For the purposes of this study, a simulated dropout was defined as a sample member who had self-reported high school grades below C, was in the lowest quartile of SES,* and was not in any kind of post-secondary education by the fall of 1974. (High school grades were used because they reflect not only an individual's academic ability but also his motivation. Likewise, SES reflects family socioeconomic condition and parental expectations. As discussed previously, these variables are highly related to dropping out of school.) The rest of the sample members, by definition, were graduates. Some of them, however, did not have information about high school grades or SES, and thus were unclassifiable. This group of members was excluded from comparisons. The final numbers of NLS participants involved in this study were 127 self-reported dropouts, 980 simulated dropouts, and 15,771 graduates.

* SES was based upon a composite of father's education, mother's education, parental income, father's occupation, and a household items index. Factor analysis revealed a common factor with approximately equal loadings for each of the five components. Missing components were imputed as the mean of the subpopulation of which the respondent was a member, defined according to cross-classifications of race, high school program, and aptitude. The available standardized components, both imputed and nonimputed, were averaged to form an SES when at least two nonimputed components were available. The continuous SES score was then assigned to one of the quartiles on the basis of the weighted frequency distribution of the composite score. The first quartile, the middle two quartiles, and the fourth quartile were respectively denoted as the low, middle, and high SES.

IV. FURTHER DESCRIPTION OF DROPOUTS AND GRADUATES

Table 1.

SOME BACKGROUND CHARACTERISTICS OF DROPOUTS AND GRADUATES

Characteristics	Self-Reported Dropouts	Simulated Dropouts	Graduates
Region			
Northeast	18.1%	16.3%	22.0%
North Central	29.9	31.7	27.1
South	23.6	39.2	32.7
West	28.3	12.8	18.3
Academic Ability*			
High	6.3	1.8	25.4
Middle	16.5	24.1	42.3
Low	50.4	65.7	25.2
Unknown	26.8	8.4	7.0
Socioeconomic Background			
High	10.2	0.0	24.9
Middle	52.0	0.0	50.7
Low	37.0	100.0	24.5
High School Program			
General	53.8	48.8	33.2
Academic	9.4	8.3	43.0
Voc-Tech	32.3	43.0	23.8
Sex			
Male	65.4	58.8	48.9
Female	34.6	41.2	51.1
Race			
Black	14.2	25.1	11.8
Hispanic	7.9	9.3	4.3
White	70.1	56.5	78.6
Other	7.8	9.1	5.3
Type of Residence Community			
Rural	23.6	30.7	21.6
Small City or Town	18.9	25.2	26.2
Large City	36.2	25.8	27.6
Suburban	15.0	13.6	22.3
Unknown	6.3	4.7	2.4
N	127	980	15,771

* The ability measure was a composite score of four tests: Vocabulary, Reading, Letter Groups, and Mathematics. A factor analysis revealed a general academic ability factor that was represented by an equally weighted linear composite of these four standardized tests. The composite score was classified into a low, middle, or high category corresponding to the first quartile, the middle two quartiles, or the fourth quartile.

Dropouts (both self-reported and simulated) and graduates were further examined with respect to their background and characteristics. As shown in Table 1, simulated dropouts have a great similarity to self-reported dropouts. Both groups had greater percentages of members from the South and North Central regions and from non-college-preparatory programs as compared to the graduate group. They also were composed of a higher proportion of non-whites and men and had lower academic ability and SES than graduates. Both self-reported and simulated dropouts were more likely to come from rural areas or large cities than suburban areas.

Prior research has shown that dropouts are generally more negative toward school than nondropouts (e.g., Bachman, Green, & Wirtanen, 1971). This was also true among the simulated and self-reported dropouts. As shown in Table 2, a substantially higher percentage of dropouts than graduates expressed such negative attitudes toward school as "don't feel part of the school" and "parents aren't interested in my education."

Simulated dropouts, like self-reported dropouts, were older than graduates. The average age levels in October 1972 were

Table 2

PERCENTAGE OF DROPOUTS AND GRADUATES EXPRESSING NEGATIVE ATTITUDES TOWARD SCHOOL

Attitude	Self-Reported Dropouts	Simulated Dropouts	Graduates
1. Plan to attain less than high school graduation.	10.31%	9.16%	1.91%
2. Courses are too hard.	8.42	4.61	1.96
3. Parents aren't interested in my education.	24.20	27.72	13.22
4. Don't feel part of the school.	17.03	14.63	9.87
5. Find it hard to adjust to school routine.	14.71	8.29	5.71
N	98	962	15,415

18.87, 18.73, and 18.50, respectively, for simulated dropouts, self-reported dropouts, and graduates. The data supported previous findings that dropouts tend to be behind grade level (Dentler & Warshauer, 1965).

In summary, the simulated dropouts had characteristics and attitudes generally observed among actual high school dropouts. The profile seemed to agree with what would be expected on the basis of previous research. However, it should be noted that the simulated as well as self-reported dropouts were seniors during the NLS base-year survey. Differences in survey data between dropouts and graduates would probably be greater than those observed in this study because most dropouts leave school before the twelfth grade.

V. ANALYSES AND RESULTS

A. Differences in Overall Participation Rates

The first question addressed in this study is: Are dropouts less likely than graduates to participate in a survey? To answer this question, the percentage of members who participated in the base year and the two follow-up surveys was computed for the dropout and the graduate groups. Results are presented in Table 3. In the base-year survey, the difference between simulated dropouts and graduates was not substantial, but the self-reported dropouts had a much lower participation rate. It is possible that those self-reported dropouts were absent or had already left school at the time the Student Questionnaire was administered in school in the spring of 1972. Thus, their lower participation rate could be due to the fact that some of them did not receive the questionnaire. In the first follow-up survey, the simulated dropouts had a lower rate than graduates, and the 100 percent participation rate of self-reported dropouts was expected because only those members who responded to the first follow-up questionnaire could have information for dropout classification. In the second follow-up survey, the self-reported dropouts and graduates had similar participation rates, but the simulated dropouts had a lower rate. The inconsistency between the self-reported dropouts and simulated dropouts could be due to varying reasons. For example, all simulated dropouts were from the lowest quartile of SES, and low SES members tended to have lower participation rates (e.g., Vincent, 1964). It is also possible that those self-reported dropouts were probably those members who were more likely to participate in the survey; otherwise, they could have dropped out of the first follow-up survey and would not be included in this group.

Table 3

PARTICIPATION RATES IN EACH SURVEY

	Base Year			First Follow-Up			Second Follow-Up		
	Self-Reported Dropouts	Simulated Dropouts	Graduates	Self-Reported Dropouts	Simulated Dropouts	Graduates	Self-Reported Dropouts	Simulated Dropouts	Graduates
Participants	77.2%	98.2%	97.7%	100.0%	87.6%	94.5%	92.9%	85.2%	93.3%
Nonparticipants	22.8	1.8	2.3	0.0	12.4	5.5	7	14.8	6.7

Note:

1. Sample n's — 127 self-reported dropouts; 980 simulated dropouts; and 15,771 graduates.
2. Self-reported dropouts were identified based upon responses to an education-status question in the first follow-up questionnaire.

Although the results did not provide a clear basis for evaluating whether or not dropouts are less likely to participate in a survey, they did show that dropouts had higher withdrawal rates. The participation rate of simulated dropouts changed from 98 percent in the base year to 85 percent in the second follow-up, while the participation rate of graduates changed from 98 to 95 percent. The self-reported dropouts also showed a 7 percent decline from the first to the second follow-up survey. (The base-year rate was not used for this comparison because the questionnaire might not have been administered to some of them because of their absence.) This may create some biases in the long run if the same participation trend continues; that is, after extensive mail and interview effort, the survey sample is comprised of fewer persons with high school dropout characteristics.

B. Differences in Mail Return Rates

Differences in the overall participation rate may be masked by the extensive efforts to interview those who failed to return questionnaires by mail. To examine this possibility and for its own value, the following question was asked: Are dropouts less likely to respond by mail if they do participate? The answer to this question is clearly shown in Table 4. Dropouts, both self-reported and simulated, had a much lower mail return rate than graduates. In the first follow-up survey, the mail return rates were about 39, 47, and 71 percent, respectively, for self-reported dropouts, simulated dropouts, and graduates. Although the mail return rates increased for each group in the second follow-up survey, substantial differences between dropouts and graduates still existed. Results clearly indicate a greater difficulty in collecting data from dropouts by mail questionnaires although they may participate in the study if interviews are conducted.

C. Differences in Tracing Efforts

Some previous studies have indicated that less educated people are easier to locate probably because of their being less mobile (e.g., Crider & Willits, 1973). The present study, however, does not support this finding. Based upon a subsample of 100 cases from each group, and considering only the first follow-up survey, dropouts in general required more tracing efforts to update their mailing addresses than graduates. As shown in Table 5, the numbers (here equal to percents) requiring telephone tracing were 28, 34, and 21, respectively, for self-reported dropouts, simulated dropouts, and graduates. Furthermore, among those members requiring telephone tracing, 39 percent (i.e., 11/28 X 100) of self-reported dropouts and 50 percent of simulated dropouts, as compared to 29 percent of graduates, had an unlisted phone or did not have a private phone, and thus they required more indirect calls. Even among those having private phones, dropouts needed more calls to complete the tracing. The average number of tracing calls was about 6.8 (i.e., 115/17) for self-reported dropouts, 6.5 for simulated dropouts, and 5.7 for graduates. In terms of telephone costs, dropouts were more expensive to trace than graduates.

Table 4

PERCENTAGES OF MAIL RETURNS AND INTERVIEWS

	First Follow-Up			Second Follow-Up		
	Self-Reported Dropouts	Simulated Dropouts	Graduates	Self-Reported Dropouts	Simulated Dropouts	Graduates
Mail Return	39.37	46.50	71.30	51.69	61.56	75.44
Interview	60.63	53.50	28.70	48.31	38.44	24.56
N	127	858	14,899	118	835	14,712

Note: No interviews were conducted in the base-year survey.

Table 5

TRACING EFFORTS FOR DROPOUTS AND GRADUATES (FIRST FOLLOW-UP)

	Self-Reported Dropouts	Simulated Dropouts	Graduates
Number of Participants Requiring Tracing	28	34	21
Having Private Phone	17	17	15
Having No Private Phone	11	17	6
Number of Participants Not Requiring Tracing	72	66	79
Total Phone Calls	190	275	128
Having Private Phone	115	110	85
Having No Private Phone	75	165	43
N	100	100	100

D. Differences in Weekly Postal Return Rates

Another question addressed in this study is. Does it take a longer time to get responses from dropouts than graduates if they do return a mailed questionnaire? To answer this question, weekly and cumulative postal return rates using date of receipt as the basis were computed for the three groups. Results are presented in Tables 6 and 7, respectively, for the first and second f It can be easily seen that there ibs.en.ial difference in postal return rates between dropouts and graduates in the first few weeks. For example, during the first week after initial mailout of the First Follow-Up Questionnaire, about 28 percent of graduates returned the questionnaire as compared to 14 percent of the simulated dropouts. Proportionally, about twice as many graduates as simulated dropouts returned the questionnaire in the first

Table 6

RELATIVE FREQUENCY OF POSTAL RETURNS BY NUMBER OF WEEKS FROM THE FIRST QUESTIONNAIRE MAILING (FIRST FOLLOW-UP)

	Frequency			Cumulative Frequency		
Weeks	Self-Reported Dropouts	Simulated Dropouts	Graduates	Self-Reported Dropouts	Simulated Dropouts	Graduates
1	7.9	13.6	28.1	7.9	13.6	28.1
2	2.4	5.4	10.6	10.3	19.0	38.7
3	0.8	2.1	4.2	11.1	21.1	42.9
4	0.8	1.3	3.3	11.9	22.4	46.2
5	2.4	1.8	2.4	14.3	24.2	48.6
6	10.2	5.0	5.1	24.5	29.2	53.7
7	5.5	3.6	4.1	30.0	32.8	57.8
8	2.4	1.5	2.4	32.4	34.3	60.2
9	0.0	1.0	1.3	32.4	35.3	61.5
10	2.4	0.6	0.8	34.8	35.9	62.3
11	2.4	0.3	0.5	37.2	36.2	62.8
12	0.0	0.4	0.2	37.2	36.6	63.0
12	2.2	4.1	4.4	39.4	40.7	67.4
N	127	980	15,771	127	980	15,771

– second questionnaire mailout.

12 – more than 12 weeks.

Table 7

RELATIVE FREQUENCY OF POSTAL RETURNS BY NUMBER OF WEEKS FROM THE FIRST QUESTIONNAIRE MAILING (SECOND FOLLOW-UP)

Weeks	Frequency			Cumulative Frequency		
	Self-Reported Dropouts	Simulated Dropouts	Graduates	Self-Reported Dropouts	Simulated Dropouts	Graduates
1	18.1	19.8	25.6	18.1	19.8	25.6
2	9.4	9.5	16.3	27.5	29.3	41.9
3	5.5	4.9	6.6	33.0	34.2	48.5
4	1.6	3.4	4.6	34.6	37.6	53.1
→ 5	1.6	1.5	2.0	36.2	39.1	55.1
6	0.0	2.9	3.2	36.2	42.0	58.3
7	2.4	2.4	3.0	38.6	44.4	61.3
8	0.8	1.9	1.2	39.4	45.3	62.5
9	0.8	0.5	0.7	40.2	45.8	63.2
10	0.8	1.2	1.2	41.0	47.0	64.4
11	0.0	1.0	0.8	41.0	48.0	65.2
12	0.0	0.3	0.4	41.0	48.3	65.6
>12	7.0	4.1	4.8	48.0	52.4	70.4
N	127	980	15,771	127	980	15,771

→ – second questionnaire mailout.

>12 – more than 12 weeks.

four weeks during the first follow-up survey. This ratio was slightly reduced in the second follow-up survey.

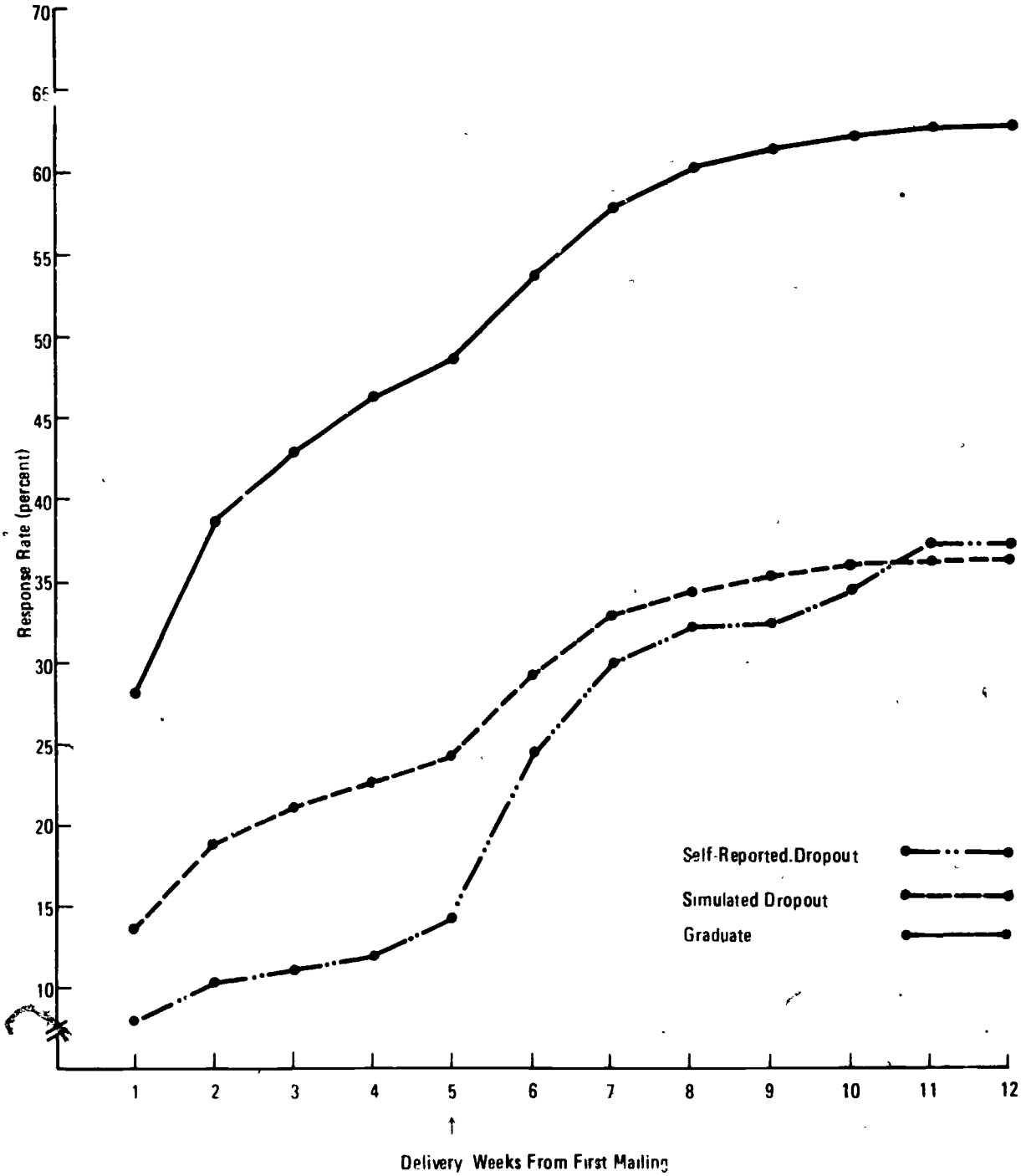
The cumulative frequency of postal returns for the groups was graphed in Figures 1 and 2, respectively, for the first and second follow-ups. These curves depict the similarity in postal return patterns between self-reported and simulated dropouts. The curves help reveal the length of time required for each group to reach a certain response rate. All groups developed their total response rates rapidly in the first few weeks, and then leveled off. A slight jump in the sixth week is due to a second mailout.

It can be seen that it took a much longer time for dropouts to reach the early graduate response rates. For example, it took about nine weeks for simulated dropouts to reach a 35 percent postal return rate whereas it took graduates less than two weeks to do so in the first follow-up survey. The return rates of dropouts after 12 weeks were still less than the percentage of questionnaires returned by graduates after only two weeks (38.7 percent).

Although postal return rates for dropouts were slightly higher in the second follow-up than in the first follow-up, the differences in questionnaire response data between dropouts and graduates were still considerable. After 12 weeks of data collection efforts, the return rates for dropouts were still less than the rate for graduates at the three-week point.

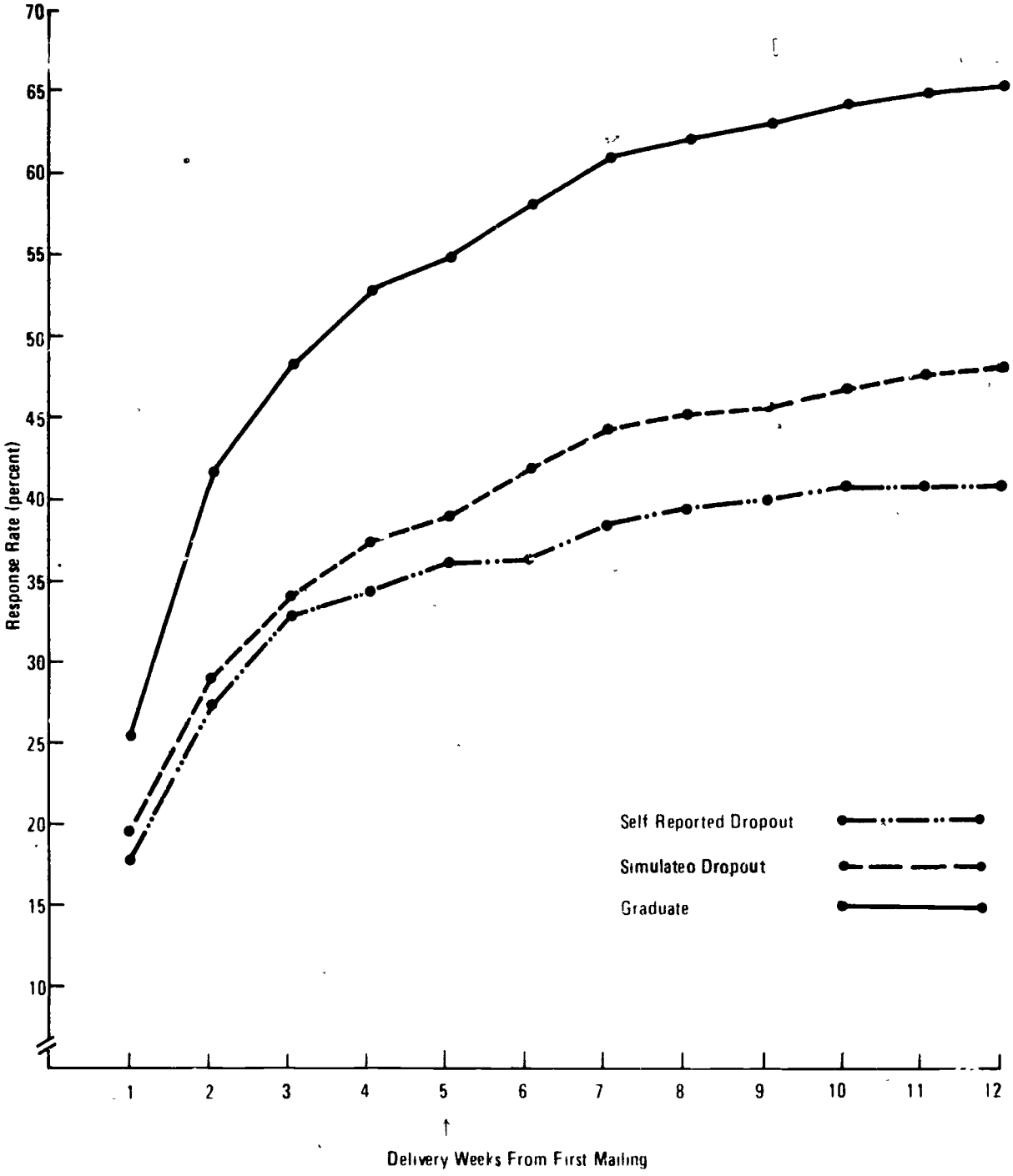
E. Differences in the Extent of Editing

Before being keyed onto tapes, all NLS questionnaire responses went through manual editing to check inconsistencies and the completeness of key questions. Any respondents who failed these edit checks were called up to clarify or complete their answers.



Note ↑ indicates the second questionnaire mailout.

Figure 1. Cumulative Postal Return Rate (First Follow-Up Survey)



Note ↑ indicates the second questionnaire mailout

Figure 2. Cumulative Postal Return Rate (Second Follow Up Survey)

Extensiveness of editing provides another basis for comparison between graduates and dropouts: Are dropouts more likely to fail edit checks than graduates?

Of those 100 randomly selected cases from each group, 39 percent of self-reported dropouts and 41 percent of simulated dropouts returned the First Follow-Up Questionnaire by mail. However, 77 of the selected graduates responded to the mailed questionnaire. Among these respondents, the percentage of dropouts who failed editing was about twice as large as that of graduates. The corresponding percentages for the three groups were 41, 42, and 22 percent. It is thus concluded that dropouts' questionnaires required a greater extent of editing (see Table 8).

Table 8

PERCENTAGE OF MAIL RETURN QUESTIONNAIRES THAT FAILED EDITS (FIRST FOLLOW-UP)

	Self-Reported Dropouts	Simulated Dropouts	Graduates
Fail Edit	40.82	42.11	22.08
Pass Edit	59.18	57.89	77.92
N	39	41	77

F. Differences in Quality Measures

It has been shown that dropouts are less likely to respond to a mailed questionnaire, take a longer time if they do respond to a questionnaire, and need more editing efforts than graduates The next question is: Do dropout and graduate participants provide data of similar quality? That is, how consistent, reasonable (e.g., not out-of-range), and complete are their responses? Do they follow directions correctly?

To answer these questions, dropouts and graduates were compared on several quality measures. The derivations of those measures are briefly described as follows:

1. Inconsistency Index (IS)

This index represents the percentage of a set of consistency checks failed by an individual. The index was computed by the following formula:

$$IS = \left(\frac{\sum_{i=1,n} X_i}{n} \right) \cdot 100$$

where $X_i = \begin{cases} 0 \text{ if the respondent passed check } i; \\ 1 \text{ if the respondent failed check } i; \end{cases}$

n = number of consistency checks.

For the first follow-up questionnaire, there were 94 consistency checks. For the second follow-up questionnaire, two inconsistency indices were developed with Index 1 calculated from 36 manual edit checks for key items, and Index 2 from 21 edit checks for non-key items.

2. Out-of-Range Index (OR)

This index represents the percentage of out-of-range responses for an individual's record and was computed as follows:

$$OR = N/D \cdot 100$$

where N = the number of items with an out-of-range code; and

D = the number of items with responses.

3. Routing Error Index (RE)

This index represents the percentage of the routing questions that were ambiguously answered by an individual (i.e., routing questions which were unanswered or answered in a manner which was inconsistent with the respondent's subsequent pattern of response). The routing error index was computed as follows:

$$RE = N/D \cdot 10($$

where N = the number of flagged routing questions; and

D = the number of routing questions with responses other than legitimate skip.

4. Incompleteness Index (IP)

For a particular section of the questionnaire, the incompleteness index represents the percentage of items with invalid responses—that is, responses that are coded as errors or missing data (first follow-up codes 93-98, second follow-up codes 94-98). The index was computed as follows:

$$IP_i = N/D \cdot 100$$

where i = the particular section of the questionnaire;

N = the number of items with invalid responses (responses with error or missing data codes); and

D = the number of items with responses other than legitimate skip.

In the following analyses, dropouts and graduates were further classified by mode of data collection (i.e., mail versus interview). This was necessary because more dropouts than nondropouts were interviewed, and interview data are assumed to be more complete than mail responses.

The analysis results, based upon the released NLS data of the dropouts and graduates, are presented in Tables 9 and 10. All the measures are in a negative direction; thus, the higher measures indicate poorer quality of the data.

The general pattern exhibited in the tables is that dropouts, both self-reported and simulated, tended to provide poorer quality data than graduates in their mail-returned questionnaires. They were more likely than graduates to provide inconsistent and incomplete responses, and to make more routing errors (i.e., not following directions correctly). In the first follow-up, the difference between simulated dropouts and graduates was largest in quality measures of routing error and incompleteness in the education and work experience sections. In the second follow-up, the difference again was very high in routing error and incompleteness of the last three sections of the questionnaire.

The magnitude of variance of quality measures, as represented by standard deviations, also indicates that dropouts were more heterogeneous than graduates with respect to data quality.

Tables 9 and 10 also present the measures of the quality of data collected through interviews. They did not show any substantial or consistent differences between dropouts and graduates. This seems to indicate that the quality of interview data is not dependent upon type of respondent.

As mentioned previously, responses to a few key items in the first and the second follow-up questionnaires were edited and corrected through phone calls if response inconsistencies occurred. This editing operation may mask some differences between dropouts and graduates. To test this assumption, unedited data obtained from the random sample of 100 participants from each of the simulated dropout and the graduate groups were analyzed. Results, as presented in the Appendix, showed that basic differences in the quality of mail-returned data did exist between dropouts and graduates. That is, if the unedited data were used in comparisons of data quality, the existence and magnitude of differences between dropouts and graduates would still be substantial; the editing operation has not masked the basic differences.

It seems reasonable to conclude that dropouts will not provide as good data as graduates through mail questionnaires. Dropouts are more likely to make routing mistakes and are less likely to complete the questions, possibly because of poorer reading skills. If a mail survey is adopted in a study, the problem of missing or inconsistent data will be more severe among dropouts, and thus may contribute some bias to analyses involving dropouts. To insure data quality, interview procedures will be preferable for dropouts.

Table 9

MEANS AND STANDARD DEVIATIONS OF DATA QUALITY INDICES
(FIRST FOLLOW-UP)

Data Quality Index	Mail			Interview		
	Self-Reported Dropouts	Simulated Dropouts	Graduates	Self-Reported Dropouts	Simulated Dropouts	Graduates
1. Inconsistency	2.30 (3.72)	1.16 (1.99)	1.26 (1.82)	0.82 (1.13)	0.69 (1.43)	0.80 (1.08)
2. Out-of-Range	0.04 (0.20)	0.14 (0.97)	0.08 (0.[illegible])	0.01 (0.11)	0.02 (0.21)	0.04 (0.38)
3. Routing Error	5.52 (8.05)	5.68 (6.39)	3.28 (5.10)	2.43 (4.64)	2.17 (4.03)	2.51 (3.67)
4. Incompleteness in Section:						
a. General Information	14.88 (9.11)	20.57 (12.78)	13.99 (25.36)	7.87 (10.14)	17.70 (26.84)	17.18 (26.53)
b. Education and Training	35.86 (35.52)	40.18 (39.67)	24.44 (26.37)	12.87 (25.74)	22.06 (37.85)	23.72 (32.63)
c. Civilian Work Experience	12.62 (22.98)	20.10 (27.78)	11.04 (20.31)	3.56 (8.65)	15.32 (30.66)	15.25 (30.56)
d. Military Service	1.26 (5.97)	3.57 (16.02)	1.47 (10.34)	3.34 (13.43)	2.90 (15.27)	2.61 (14.79)
e. Information About the Past	10.94 (21.43)	13.58 (24.08)	8.00 (16.94)	3.96 (12.74)	16.66 (35.24)	18.31 (34.78)
N	50	399	10,623	77	459	4,276

Note: Figures in parentheses are standard deviations.

Table 10

MEANS AND STANDARD DEVIATIONS OF DATA QUALITY INDICES
(SECOND FOLLOW-UP)

Data Quality Index	Mail			Interview		
	Self-Reported Dropouts	Simulated Dropouts	Graduates	Self-Reported Dropouts	Simulated Dropouts	Graduates
1. Inconsistency (Key Items)	0.90 (2.17)	0.46 (1.67)	0.29 (1.02)	0.07 (0.37)	0.12 (0.64)	0.17 (0.69)
2. Inconsistency (Non-key Items)	2.62 (4.64)	1.90 (3.61)	1.73 (3.46)	2.82 (4.28)	2.61 (4.45)	3.31 (5.97)
3. Routing Error	6.02 (8.67)	3.85 (6.72)	2.18 (3.79)	2.84 (4.42)	2.27 (3.08)	2.09 (2.98)
4. Out-of-Range	0.02 (0.13)	0.02 (0.10)	0.02 (0.38)	0.00 (0.09)	0.01 (0.10)	0.01 (0.14)
5 Incompleteness in Section:						
a. General Information	2.66 (7.22)	2.18 (7.58)	1.31 (5.91)	12.70 (27.19)	18.06 (31.47)	14.95 (29.23)
b. Education and Training Training	27.69 (35.39)	20.38 (37.06)	15.85 (24.26)	18.75 (33.72)	24.73 (37.06)	24.13 (34.72)
c. Work Experience	20.43 (27.00)	17.08 (24.23)	14.63 (21.63)	24.07 (35.11)	27.83 (38.78)	24.36 (36.11)
d. Family Status	11.69 (12.44)	11.38 (15.88)	7.25 (11.48)	19.89 (37.29)	25.74 (41.53)	21.74 (39.00)
e. Military Service	5.93 (20.38)	3.90 (16.14)	1.25 (8.66)	2.68 (13.00)	5.62 (19.16)	2.76 (13.94)
f. Activities and Opinions	8.48 (13.09)	5.91 (12.04)	3.32 (7.61)	20.33 (39.18)	25.45 (42.51)	21.77 (39.85)
N	61	514	11,099	57	321	3,613

Note: 1. Figures in parentheses are standard deviations.

2. The comparison between interview and mail responses is not legitimate because some questions were not asked in interviews due to the nature of the questions.

VI. SUMMARY AND DISCUSSION

The primary purpose of this study was to examine the problems in data collection from high school dropouts. A group of self-reported dropouts and a group of simulated dropouts were compared with graduates on their participation rate, tracing efforts, postal return delivery date, and the quality of their responses. Major findings are summarized in the following paragraphs.

Postal return rates are lower among dropouts. The difference between dropouts and graduates generally runs between 25 percentage points in the first follow-up and 18 percentage points in the second follow-up. To maintain a high participation rate, dropouts will require more phone or field interviews than graduates. Dropouts also need more tracing efforts to update their mailing address, partly due to their greater tendency to not return mailed materials.

Dropouts are less likely than graduates to respond to a mail survey promptly. At the end of 12 weeks the return rates of dropouts were no higher than those of graduates after only two or three weeks.

About 40 percent of the dropouts' questionnaires failed editing and required further phone calls to complete or correct their questionnaires. This was about twice as many as graduates. In addition, the overall quality of the data from dropouts is less satisfactory than that from graduates. dropouts tend to provide more erroneous responses (e.g., out-of-range, inconsistent, and incomplete responses) than graduates.

It is reasonable to conclude that it is much more difficult to obtain quality data by mail from dropouts. In addition, the declining participation rate of dropouts through time may bias the sample in later follow-ups. If the dropout component is included in a sample, additional efforts should therefore be taken to maintain the dropout participation and at the same time insure high quality data.

The interview approach, either through phone or field interviews, is preferable for dropouts with respect to maintaining a high response rate and high quality of data. However, a combination of mail survey and interview may prove to be cost effective. Mail return may be improved by making the questionnaire as simple and straightforward as possible. Many dropouts may have difficulty in reading, and thus difficult language, complex skip patterns, and a lengthy questionnaire may confuse or discourage many dropouts from participating. To the extent possible, these barriers should be removed. In addition, mail return may be increased by some incentive mechanism. The offer of a $3.00 incentive in the third follow-up survey will present an opportunity for testing the influence of incentives on dropouts and graduates when the data are available. It seems that the motivation (willingness) to participate and the ability (i.e., reading skill) to complete a questionnaire are the two major factors in a mail survey. Steps to cope with these two major factors are advisable in future studies involving dropout components.

References

Bachman, J. G., Green, S., & Wirtanen, I. D. *Youth in transition. Vol. 3 - Dropping out—problem or symptom.* Ann Arbor, Michigan: The University of Michigan, 1971.

Crider, D. M., & Willits, F. K. Respondent retrieval bias in a longitudinal survey. *Sociology and Social Research*, 1973, *58*(1), 56-65.

Dentler, R. S., & Warshauer, M. E. *Big city dropouts and illiterates.* New York: Frederick A. Praeger, 1968.

Lenski, G. E., & Leggett, J. C. Caste, class, and difference in the research interview. *The American Journal of Sociology*, 1960, *65*(5), 463-467.

Moser, C. A., & Kalton, G. *Survey methods in social investigation.* New York: Basic Books, Inc., 1972.

Nuess, C. F. Differences between persons responding and not responding to a mailed questionnaire. *American Sociological Review*, 1943, *8*(4), 433-438.

U.S. Bureau of Census. Census of population: 1970, subject reports, final report pc(2)-5A, *School Enrollment*, 1973.

Vincent, C. E. Socioeconomic status and familial variables in mail questionnaire responses. *The American Journal of Sociology*, 1964, *64*(6), 647-653.

Appendix

Means and Standard Deviations of Data Quality Indices

Appendix

MEANS AND STANDARD DEVIATIONS OF DATA QUALITY INDICES (UNEDITED FIRST FOLLOW-UP DATA FROM RANDOMLY DRAWN SAMPLES)

Data Quality Index	Mail		Interview	
	Simulated Dropouts	Graduates	Simulated Dropouts	Graduates
Out-of-Range	0.03	0.04	0.00	0.00
	(0.16)	(0.22)	(0.00)	(0.00)
Routing Error	5.16	3.51	2.30	2.47
	(4.86)	(5.52)	(4.39)	(3.22)
Incompleteness in Section:				
a. General Information	19.57	15.09	14.70	13.71
	(11.79)	(10.48)	(25.09)	(24.92)
b. Education and Training	48.92	21.87	15.09	19.76
	(38.67)	(22.20)	(32.13)	(30.02)
c. Civilian Work Experience	26.05	10.87	12.25	9.57
	(29.50)	(18.58)	(26.08)	(24.53)
d. Military Service	2.84	2.86	1.93	0.00
	(11.79)	(15.82)	(11.76)	(0.00)
e. Information About the Past	12.05	8.25	13.91	16.86
	(19.36)	(19.51)	(33.04)	(30.73)
N	37	77	44	21

Note: No inconsistency index was computed because of the complexity of computer program changes required and the insignificant number of changes in responses to the items involved in the computation.