DOCUMENT RESUME

ED 217 791 HE 015 187

TITLE	Quality in the Basic Grant Delivery System: Volume 3, Methodology.
INSTITUTION	Advanced Technology, Inc., McLean, Va.; Westat Research, Inc., Rockville, Md.
SPONS AGENCY	Office of Student Financial Assistance (ED), Washington, DC.
PUB DATE	Apr 82
CONTRACT	300-80-0952
NOTE	154p.; For related documents, see HE 015 184-186.
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Accountability; College Students; Confidential Records; Data Collection; Disclosure; Eligibility; Error Patterns; Family Financial Resources; Family Income; *Federal Aid; Financial Aid Applicants; Financial Needs; Grants; Money Management; *Need Analysis (Student Financial Aid); Parent Financial Contribution; Postsecondary Education; Program Administration; *Quality Control; *Research Methodology; *Student Financial Aid
IDENTIFIERS	*Basic Educational Opportunity Grants; Pell Grant Program

ABSTRACT

The research methodology of a study to assess 1980-1981 award accuracy of the Basic Educational Opportunity Grants (BEOG), or Pell grants, is described. The study is the first stage of a three-stage quality control project. During the spring of 1981 a nationally representative sample of 305 public, private, and proprietary institutions was visited. The financial aid administrator at each institution was interviewed and asked to describe the institution's BEOG processing procedure. At each institution, data from a random sample of an average of 14 financial aid records were reviewed and transcribed. In all, data were collected from 4,500 BEOG recipient records. These recipients and their parents were interviewed, asked about their general experiences in dealing with the application process, and asked to provide documents to verify the income and household information on their application forms. In addition, data were collected from the Internal Revenue Service, tax assessors, and financial institutions as additional verification of the information that the students in the survey placed on their applications. Information is presented concerning the rationale and specific procedures used to select a statistically representative sample of institutions and students; the response rates for the survey of students and parents; institutional and student/parent field data collection; and procedures used to compile, edit, and convert the survey data to machine-readable tapes used for statistical analysis. Attention is also directed to the methodology and procedures used in the special analysis of BEOG application processor data entry error. (SW)

QUALITY IN THE BASIC GRANT DELIVERY SYSTEM

Volume 3
Methodology

Submitted to
Office of Student Financial Assistance
Department of Education

Contract No. 300-80-0952

April 1982

Advanced Technology, Inc.
7923 Jones Branch Drive
McLean, Virginia 22102

and

Westat, Inc.
1650 Research Boulevard
Rockville, Maryland 20850

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

CHAPTER 1

INTRODUCTION

PROJECT OVERVIEW

The Office of Student Financial Assistance [OSFA] of the Department of Education has contracted with Advanced Technology, Inc. of McLean, Virginia, and Westat, Inc. of Rockville, Maryland, to conduct a three-year quality control study of the Basic Educational Opportunity Grant [BEOG] Program. The goal is to assess the accuracy and reliability of the BEOG program and improve its administration.

The project is being conducted in three stages. During the first year of the project--or Stage One--the objectives have been to measure program error rates, to determine the cause of error, and to recommend changes to reduce error. Specifically, the Stage One study has:

- Determined program-wide discrepancy rates and probable causes attributable to (1) institutions, (2) recipients, (3) parents, and (4) application processors
- Identified "error prone" subpopulations among recipients
- Developed cost-benefit analyses for feasible corrective management activities to reduce error rates for every area in which error rates are excessive

In Stage Two, running from October 1981 to December 1982, an ongoing Quality Control System will be designed and tested to continuously measure and analyze BEOG program performance. In Stage Three, running from October 1982 to December 1983, the Quality Control System will be installed.

REPORT OVERVIEW

This is a report on the methodology used to accomplish the Stage One objectives just given. During the spring of 1981 a nationally representative sample of 305 public, private, and proprietary institutions was visited. The Financial Aid Administrator [FAA] at each institution was interviewed and asked to describe the institution's BEOG processing procedure. At each institution, data from a random sample of an average of 14 financial aid records were reviewed and transcribed. In all, data were collected from 4,500 BEOG recipient records. These 4,500 BEOG recipients--and their parents--were interviewed, asked about their general experiences in dealing with the application process, and asked to provide documents to verify the income and household information on their application forms. In addition, data were collected from the IRS, tax assessors, and financial institutions as additional verification of the information that the students in the survey placed on their applications. In sum, eight data sets from the following sources were collected and analyzed to meet the Stage One objectives:

- FAA interviews
- Financial aid records
- Student interviews
- Parent interviews
- Student Eligibility Reports [SER]

- IRS records
- Tax assessor records
- Financial institution records

Chapter 2 of this report describes the rationale and the specific procedures used to select a statistically representative sample of institutions and students. In addition, this chapter discusses the response rates for the survey of students and parents and analyzes the nonresponse to determine the possible nonresponse bias in estimating program-wide error amounts. Chapter 3 focuses on the institutional and student/parent field data collection. Included is a discussion of the field organization, the quality control procedures, and the procedures used to conduct the interviews. Chapter 4 describes the procedures used to compile, edit, and convert the survey data to machine-readable tapes used for statistical analysis. Chapter 5 provides a brief discussion of the methodology and procedures used in the special analysis of BEOG application processor data entry error.

The following summarizes, by chapter, the general features of the Stage One methodology.

Chapter 2--Sample Selection and Sampling Error

- In designing the sample, the goal was to provide as precise estimates as possible of the universe under study and at the same time:
  - Limit the amount of field travel
  - Control the number of student and separate institutions selected
  - Ensure that a variety of types of institutions were represented in the sample

- To accomplish this, institutions and students were selected in two stages: (1) stratification and random selection of 305 institutions and (2) stratification and random selection of approximately 4,500 BEOG recipients from these institutions. Both the institution and student sample sizes were large enough to ensure the required statistical precision.

- Institutions were stratified into four groups:

  - Institutions with over 5,000 BEOG eligible students
  - Institutions using the Regular Disbursement System RDS] and located in the 25 largest Standard Metropolitan Statistical Areas [SMSAs]
  - Institutions using RDS and located outside the 25 largest SMSAs
  - Institutions using the Alternate Disbursement System [ADS]

  All institutions in the first group were included in the institutional sample. Institutions within the other three were clustered geographically to limit travel costs and further stratified by institutional type and control (i.e., public, private, and proprietary). Institutions were selected from each substratum with probabilities proportional to size.

- Recipients from each of the 305 selected institutions were stratified by grant amount to provide better estimates of the total absolute dollar error. It was necessary to sample "validated" recipients at a higher rate than unvalidated recipients in order to provide more accurate estimates of this analytically important subgroup.

- Over 90 percent of all students and parents in the sample responded with an interview. This composite response rate can be broken down for dependent and independent students and parents as follows:

- Dependents:	students:	94.6%
	parents:	93.6%
- Independents:	students:	87.4%
	parents:	61.0%

- The analysis thus far shows that nonresponse in the survey will not significantly undermine the study's overall findings. Preliminary analysis shows that nonrespondents do not differ significantly from respondents in terms of a variety of characteristics.

Chapter 3--Data Collection and Quality Control

- Responsibility for data collection was divided between Westat and Advanced Technology. Westat conducted nationwide interviews with students and parents and collected secondary data from the IRS, tax assessors, and financial institutions. Westat's data collection occurred from February to April 1981. Advanced Technology collected institutional data. Site visits to institutions occurred from March to May 1981.

- Data collection questionnaires were developed from those used in the 1978-79 Quality Control study. In order to ensure their effectiveness, they were pilot tested in the Washington, D.C., area among a small representative sample of students, parents, and institutions.

- Westat used its nationwide network to recruit approximately 200 interviewers to conduct the student and parent interviews. Nearly all had previous survey research interviewing experience; most had worked with Westat on previous studies. All student/parent interviewers underwent rigorous training in interviewing techniques and field procedure.

- Advanced Technology recruited 13 individuals to collect the institutional data. All had considerable student financial aid experience; several had advanced degrees. It was felt that well-educated persons with recent work experience in financial aid offices were needed to conduct the specialized interviews with Financial Aid Administrators. The institutional interviewers were provided with one week of intensive training in the use of both the institutional questionnaire and student record abstract form.

- Several procedures were instituted to guarantee the timeliness and integrity of the data collection. Student/parent interviewers were managed by seven full-time in-field supervisors. These regional supervisors reported each week to the home office by telephone and

by using a computer assisted management system. The institutional interviewers were supervised directly by the home office and were required to report each week by telephone. For both survey operations, updates to procedures were communicated to interviewers by frequent field memoranda. The following were procedures used to ensure the reliability of the data:

- Each questionnaire was edited for omissions, ambiguities, or misplaced codes by the interviewer, by the field supervisor, and finally by the coding staff in the home office.
- A random sample of approximately 10 percent of all completed cases was validated. Students, parents, and Financial Aid Administrators were telephoned, asked to comment on the conduct of the interviewer, and re-asked two or three key questions.
- All interviewers were monitored in the field by senior staff at least once during the data collection.

Chapter 4--Data Processing

- All survey questionnaires and secondary data were compiled, edited, and converted to machine-readable tapes using standard ADP procedures. Standard procedures included:
  - Preliminary manual edit to check for completeness and consistency
  - Coding for data entry
  - Another manual edit for coding and interviewer errors
  - Key entry and key verification
  - Machine edit for omissions, consistency, and valid data ranges
  - Frequency distributions of each variable to check for data errors not detected by manual and machine edits

- When significant errors or omissions were discovered during the manual or machine edits, the respondent or interviewer was telephoned to retrieve the correct data.

- Once edited and entered onto tape, the clean data files were reformatted and merged into a master file for access by SAS, a statistical software package.

Chapter 5--Application Processor Data Entry Error Analysis

- A random subsample of 1,250 BEOG applications was drawn from our sample to analyze the error rate associated with data entry. All subsampled applications originated from the Multiple Data Entry [MDE] processors. The disposition of the subsample, by processor, follows:

  500 - College Scholarship Service [CSS]

  500 - American College Testing [ACT] Program

  250 - Pennsylvania Higher Education Assistance Agency [PHEAA]

- The original application forms were visually compared with the data contained on the 1980-81 Central Processor's History/Correction File. Discrepancies were tabulated manually.

CHAPTER 2

SAMPLE SELECTION AND SAMPLING ERROR

SAMPLE DESIGN AND SAMPLING PROCEDURES

Design Objectives

The primary objective of the sample design was the selection of a probability sample of students enrolled at educational institutions involved in the 1980-81 BEOG program. The statistically representative sample was used to document, compute, and analyze the program-wide error rate in program determinations including eligibility, benefit amount, and payments. To do this, data were obtained from students, their parents, and educational institutions.

In terms of survey implementation, the sample design had the following objectives:

- Limit the amount and cost of field travel
- Control the number of students and separate institutions selected
- Ensure the representation of a variety of institution types
- Provide as precise estimates as possible

The procedures described in the following, which employs selection of units with probabilities proportionate to size, yielded a sample meeting these objectives. Development of estimated measures of size is discussed in the following paragraphs.

An important goal in sample design is to organize available information for drawing the sample so as to reduce the variability of the characteristics of possible samples or to increase

the precision of estimates. Ideally, the sample design would result in a set of possible samples each of which closely resembles the universe under study, while assuring all units a known chance of selection.

Measure of Size

In an earlier survey of BEOG recipients, the number of students eligible for a grant at an institution was used to assign a selection probability for the institution. The estimated figures were used to obtain a "measure of size" [MOS] for the first-stage selection based on probabilities proportionate to size. When the MOS is not well correlated with the statistics being estimated, one of two conditions will result: either the number of students in the sample will be subject to greater variations than desired or great disparities in selection probabilities for students will occur. When the latter happens, substantial inefficiencies or large sampling errors can result.

The institution-reported number of eligibles was obtained from Department of Education data files. It is not surprising that these figures rarely matched the actual number of recipients at the time the institution was contacted. Frequently, this was due to the fact that eligible students do not always apply for grants each semester. Thus, the number of eligibiles generally overstated the size of the universe at the time of sampling. In the selection scheme described in the following, the estimated number of eligibles was used to improve the design in a manner

somewhat different from that of previous studies. This method is less affected by the accuracy of the estimates.

Size of institution was involved at two points in the design. First, the number of eligibles, indicated on the OE data tape, was used in the formation of geographic clusters of institutions. These clusters were constructed so as to contain roughly equal numbers of eligibles. For this purpose, accuracy of the size measure was relatively less critical than in the previous study.

After clusters were selected, the schools in the clusters were contacted by telephone. This fairly inexpensive effort was aimed at obtaining more precise and accurate estimates of the actual number of recipients at each institution. With such measures, a subsample of institutions was designated which allowed for greater control of student sample size and of sampling errors.

Desired Sample Sizes

One study objective was to ensure that the confidence interval for national estimates of the case and dollar error for program-wide, student, and school groups not exceed plus or minus 5 percent with 95 percent confidence. Thus, the coefficient of variation (standard deviation divided by the mean) should be less than .025 for the overall estimate of the case error rates and dollar error amounts. Below we discuss the implication of this requirement on the number of students required in the sample.

During the previous study of 1978-79 recipients, it was estimated that approximately 84 percent of all institutions had an average net dollar error of practically zero (less than $2). To estimate this characteristic with the required precision, a simple random sample of 305 institutions would be required. That is if the square of the coefficient of variation is expressed as:

$$V_p^2 = \frac{(1-p)}{p} \cdot \frac{1}{n}$$

This formula can be solved for desired sample size:

$$n = \frac{(1-p)}{p} \cdot \frac{1}{V_p^2}$$

If p, the proportion of interest, equals .84, then the number of schools required to restrict the relative variance to $(.025)^2$ is:

$$n = \frac{.16}{(.84)(.00625)} = 305$$

This characteristic is just one of many statistics of interest; therefore, use of this sample size will not produce equally accurate estimates for other characteristics.

The most important characteristics to be estimated pertained to students, not to institutions. Statements about the percentage of students with particular characteristics are more germane to the study than observations about institutions. Below we discuss the number of students desired in the sample.

An important student characteristic is that of absolute dollar error. One estimate of the dollar error can be modeled simply as the product of two random variables:

$$\bar{x} = P\bar{e}$$

where

$\bar{x}$ = average absolute dollar value of errors

P = the probability of making an error

$\bar{e}$ = the average of the absolute dollar value of the error given that an error was made

The coefficient of variation of such an estimate based on a simple random sample of n units is expressed as:

$$V_{\bar{x}}^2 = \frac{V_e^2 + Q}{nP}$$

where

$Q = 1 - p$

V_e = the coefficient of variation of the amount of error for those cases which are in error

n = the number of units (schools or students)

This formula can be rearranged to develop an expression for desired sample size:

$$n = \frac{V_e^2 + Q}{P(V_{\bar{x}}^2)}$$

Determination of the number of students required in the sample to yield the estimated coefficient of variation requires estimates

of V_e^2 and P. Based upon our previous study, $V_e^2 = .935$ and P = .431. Using these values, a sample of 5,600 students would be needed:

$$n = \frac{.935 + (1-.431)}{(.431)(.025)^2} = 5583$$

We felt that the value of V_e^2 in this study would be considerably below .935 for two reasons: (1) The method of selecting clusters differed from the prior study in a way that reduced homogeneity within the clusters. This resulted in smaller between-cluster variances and in reduced design effects. Sample clusters would be created by intentionally combining schools of different sizes and hopefully reducing the rate of homogeneity (ROH). (2) The average cluster size would be reduced. By forming smaller clusters of roughly equal size and by sampling more of these clusters, as discussed below, any undesirable effects of clustering would be reduced still further.

The design effect is given by:

$$deff = 1 + (ROH)(B-1)$$

With the rate of homogeneity (ROH) and average cluster size (B) both decreasing, we expected a considerable reduction in V_e^2. From our experience, it appeared that V_e^2 should be lowered to the 0.4 to 0.5 range. At 0.5 the required sample size was about 4,000 students:

$$n = \frac{.5 + (1-.431)}{(.431)(.025)^2} = 3968$$

Institutional Sample Frame

The sample frame used for selection of institutions is the Institutional Master File [IMF] maintained by the Office of Student Financial Assistance. This file is created each year from the institutions approved for participation, and data are added as the grant program is implemented and quarterly reports are received. The 1979-80 file was used. The file identified the most up-to-date list of participating institutions and the number and value of grants awarded in the previous academic year.

For sampling purposes, it was desirable to have a comprehensive listing of schools participating in the BEOG program and to have, for each of these schools, the number of grants given for the academic year 1980-81 at the time of the survey. This latter figure served ideally as a measure of size in drawing a sample of schools. It is believed that the participating schools will be adequately identified by the 1979-80 IMF, as described above. The number of grants given in the previous year was used as proxy for the 1980-81 grants where such information was available on the IMF.

Use of previous year grants resulted in several problems. First, the institutional sample needed to select branch campuses of multicampus institutions independently to avoid excess travel costs. However, the number of grants awarded to a multicampus institution is usually included in the central campus record; blanks are recorded for number of grants to the branch campuses.

Therefore, it was necessary to allocate the total number of grants among the central institution and its branches.

If enrollments for the main campus and the branches were available, these enrollment data were used to allocate the total number of grants to the branches. If enrollment data were not available, the number of grants was distributed using other assumptions. As in the previous survey, we assumed that the main campus had three times the number of grants of any branch and that all branches had an equal number of grants.

Recipient Sample Frame

The recipient sample frame was institution records of grant recipients as of the early part of 1980-81 academic year. This frame was expected to contain virtually all 1980-81 grant year recipients. However, certain types of students were omitted such as late grant applicants, students whose applications have not been approved, students enrolling for second term, etc.

Sample Selection Procedures

The sampling procedures involved two phases: (1) selection of a sample of institutions and (2) selection of a sampling of students at the selected institutions.

Selection of Institutions

The first step in selection of the institution sample was the sorting of schools from the master file into four groups:

- Institutions with over 5,000 eligibles
- Institutions using Regular Disbursement System [RDS] located in the 25 largest SMSAs

- Institutions using Regular Disbursement System [RDS] not located in the 25 largest SMSAs
- Institutions using Alternative Delivery System [ADS] or institutions for which there were no available eligibility counts on the master file

The sample size was allocated to these four groups according to the number of eligibles. For the fourth group, eligibility counts were estimated using data from the earlier BEOGs study. All institutions in the first group were included in the institution sample. Institutions were selected from the other three groups as described in the following paragraphs.

Institutions Using RDS in the 25 Largest SMSAs

Sample selection involved five steps for this group:

1. Select a systematic sample of about 150 institutions from a list of all schools ordered by SMSA, type of control, type of school, and number of eligibles. Schools with large numbers of eligibiles, i.e., where the number of eligibiles exceeded the sampling interval, were included with certainty at this stage.
2. Contact the schools selected above and obtain an improved measure of size i.e., number of recipients. This new measure of size will be used in the final probability proportional to size selection stage.
3. Form substrata using type of control, type of institution, and size (eligibility counts).
4. Allocate sample to substratum according to new measure of size (recipients).
5. Select desired sample from each substratum with probability proportional to size.

Institutions Using RDS Not Located in 25 Largest SMSAs

Institutions in this group were selected using an eight-step procedure. Geographic clustering was utilized for this group in order to control interviewing costs.

1. Sort institutions according to geographical location using Westat's Zip Recode.

2. Form geographic clusters of nearly equal size. Nearly equal size means that the smallest cluster in terms of the number of eligibles cannot be less than one half the average size nor greater than twice the average size. Clusters formed by grouping multiple institutions most often yielded a work load of 20 to 29 cases; single institution clusters most often produced 30 to 39 cases.

3. Combine the above clusters into 40 primary strata by counting off one-fortieth of the clusters.

4. Select two clusters per stratum using equal probabilities without replacement.

5. Contact schools in the 80 clusters and secure a new measure of size, number of recipients. These measures were collected using telephone calls.

6. Stratify institutions according to type of control, type of institution, and size (number of eligibles).

7. Allocate sample to substrata according to new measure of size, number of recipients.

8. Select institutions from each substratum with probabilities proportional to size.

Institutions Using ADS or Institutions Where Number of Eligibles Is Not Known

Institutions in this group were selected using a two-step procedure.

1. Sort by whether institutions were inside or outside the 25 largest SMSAs, type of control, and type of institutions.

2. Select institutions from the entire list using systematic sampling with equal probabilities.

The results of these procedures are summarized in the following three tables. The first two tables cover the certainty groups (rows with size of institutions over 5,000) and the two RDS groups. The third table includes ADS schools and schools for which we had no size estimates.

Selection of Students from Sampled Institutions

Lists of students receiving Basic Grants were obtained from the sampled institutions during a period of eight weeks. If the lists were not received in the mail on a timely basis, on-side visits were made to select a sample. Students were stratified by dollar grant amounts to provide better estimates of the total absolute dollar errors.

In order to assure that at least 1,000 students identified by the institutions as "validated" students were included in the sample, we asked each institution to identify the total number of students to be validated from their institution. Then, depending on the number of total students to be validated, we separated the two lists and selected proportionately from the two lists of recipients. This provided us with at least 1,000 "validated" students.

Detailed Procedure

1. To ensure accurate probabilities in student selection, separate lists sorted by grant amount were compiled for each validation status. In smaller institutions all recipients'

Size of Institution	Institutional Universe Inside 25 Largest SMSAs								Institutional Sample Inside 25 Largest SMSAs							
	Universities 4 + years		2 years		Other		Total		Universities 4 + years		2 years		Other		Total	
	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants
Public																
0-99	7	588	16	548	9	466	32	1,602	-	---	--	--	--	--	--	--
100-	19	5,344	83	23,637	9	1,851	111	30,832	3	855	3	828	--	--	6	1,683
500-	33	31,835	89	75,637	4	2,951	126	110,423	1	752	8	6,073	--	--	9	6,825
1500-	62	173,157	44	128,402	--	--	106	301,559	8	27,534	7	22,994	--	--	15	50,528
>5k	6	44,805	2	15,106	--	--	8	59,911	6	44,805	2	15,106	--	--	8	59,911
Total	127		234		22		383		18		20		--		38	
Private																
0-99	123	5,975	107	4,821	19	538	249	11,334	1	96	1	25	--	--	2	121
100-	208	56,106	35	7,858	8	2,103	251	66,067	18	5,770	2	564	--	--	20	6,334
500-	95	79,576	6	5,190	--	--	101	84,766	7	7,102	1	576	--	--	8	7,678
1500-	27	64,833	1	3,260	--	--	28	68,093	--	--	--	--	--	--	--	--
>5k	1	5,067	--	--	--	--	1	5,067	1	5,067	--	--	--	--	1	5,067
Total	454		149		27		630		27		4		--		31	
Proprietary																
0-99	9	398	70	3,484	451	19,068	530	22,950	--	--	1	79	6	294	7	373
100-	4	1,213	77	18,353	189	40,934	270	60,500	--	--	3	654	12	3,005	15	3,659
500-	5	5,205	12	10,723	31	25,346	48	41,274	--	--	2	1,378	2	1,416	4	2,794
1500-	--	--	--	--	2	3,078	2	3,078	--	--	--	--	--	--	--	--
>5k	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Total	18		159		673		850		--		6		20		26	
Total																
0-99	139	6,961	193	8,853	479	20,072	811	35,886	1	96	2	104	6	294	9	494
100-	231	62,663	195	49,848	206	44,888	632	157,399	21	6,625	8	2,046	12	3,005	41	11,676
500-	133	116,616	107	91,550	35	28,297	275	236,463	8	7,854	11	8,027	2	1,416	21	17,297
1500-	89	237,990	45	131,662	2	3,078	136	372,730	8	27,534	7	22,994	--	--	15	50,528
>5k	7	49,872	2	15,106	--	--	9	64,978	7	49,872	2	15,106	--	--	9	64,978
Total	599		542		722		1,863		45		30		20		95	

FIGURE 2-1

INSTITUTIONAL UNIVERSE AND SAMPLE:
RDS INSIDE 25 LARGEST SMSAs

Size of Institution	Institutional Universe Outside 25 Largest SMSAs								Institutional Sample Outside 25 Largest SMSAs							
	Universities 4 + years		2 years		Other		Total		Universities 4 + years		2 years		Other		Total	
	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants	Schools	Grants
Public																
0-99	16	1,183	11	3,473	73	3,202	100	7,858	--	--	1	36	--	--	1	36
100-	72	24,708	422	115,357	24	6,280	518	146,345	4	1,166	23	7,746	--	--	27	8,912
500-	195	184,147	225	181,414	4	3,116	424	368,677	20	18,259	21	18,517	--	--	41	36,776
1500-	198	507,265	50	102,931	--	--	248	610,196	26	73,063	3	7,056	--	--	29	80,119
>5k	10	59,727	1	5,471	--	--	11	56,198	10	59,727	1	5,471	--	--	11	65,198
Total	491		709		101		1,301		60		49		--		109	
Private																
0-99	73	4,059	176	7,678	24	878	273	12,615	--	--	2	81	--	--	2	81
100-	406	120,498	92	20,330	3	986	501	141,814	22	7,630	6	1,762	--	--	28	9,392
500-	185	145,909	13	8,695	--	--	198	154,604	19	16,406	1	677	--	--	20	17,083
1500-	10	21,510	--	--	--	--	10	21,510	2	5,330	--	--	--	--	2	5,330
>5k	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Total	674		281		27		982		43		9		--		52	
Proprietary																
0-99	14	535	107	5,118	736	28,269	857	33,922	--	--	1	98	6	268	7	366
100-	7	1,495	85	20,114	174	36,428	266	58,037	--	--	5	1,483	1	183	6	1,666
500-	5	4,753	16	11,496	16	11,711	37	27,960	--	--	--	--	2	1,884	2	1,884
1500-	--	--	1	2,188	1	2,057	2	4,245	--	--	--	--	--	--	--	--
>5k	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Total	26		209		927		1,162		--		6		9		15	
Total																
0-99	103	5,777	294	16,269	833	32,349	1,230	54,395	--	--	4	215	6	268	10	483
100-	485	146,701	599	155,801	201	43,694	1,285	346,196	26	8,796	34	10,991	1	183	61	19,970
500-	385	334,809	254	201,605	20	14,827	659	551,241	39	34,665	22	19,194	2	1,884	63	55,743
1500-	208	528,775	51	105,119	1	2,057	260	635,951	28	78,393	3	7,056	--	--	31	85,449
>5k	10	59,727	1	5,471	--	--	11	65,198	10	59,727	1	5,471	--	--	11	65,198
Total	1,191		1,199		1,055		3,445		103		64		9		176	

FIGURE 2-2

INSTITUTIONAL UNIVERSE AND SAMPLE:
RDS OUTSIDE 25 LARGEST SMSAs

		Institutional Universe				Institutional Sample				
		Universities 4-yr.	2-yr.	Other	Total		Universities 4-yr.	2-yr.	Other	Total
Inside 25 Largest SMSAs	Public	9	32	31	72	Public	0	0	1	1
	Private	72	124	76	272	Private	5	4	3	12
	Proprietary	12	58	129	199	Proprietary	0	3	8	11
	Other	0	0	10	10	Other	0	0	0	0
	Total	93	214	246	553	Total	5	7	12	24
Outside 25 Largest SMSAs	Public	8	129	86	223	Public	1	5	4	10
	Private	41	188	143	372	Private	3	9	5	17
	Proprietary	7	118	221	346	Proprietary	0	5	11	16
	Other	0	0	7	7	Other	0	0	0	0
	Total	56	435	457	948	Total	4	19	20	43

FIGURE 2-3

INSTITUTIONAL UNIVERSE AND SAMPLE:
ADS STRATA

names and addresses were written on the appropriate sheet, thus effectively stratifying students by grant amount. For larger institutions, particularly when sampling on-site, a sample of 1 in k students was selected, where k is chosen to yield no more than 300 names. Only the sampled names were stratified by validation status and sorted by amount.

2. After the listing sheets had been completed, a systematic sample was drawn from each list. It was necessary to sample validated students at a higher rate. The sampling fractions were established so that the final samples would be self-weighting within the two groups, validated and unvalidated. It was necessary to sample validated students at a higher rate to achieve the desired sample size.

3. Based on the desired 4,000 completed student instruments, 1,000 of whom were validated from 305 participating institutions, and an anticipated completion rate of .87 (this was the completion rate for the 1978-79 study), an average of 15 recipients per institution were selected.

While the two samples are self-weighting, the sample weights for unvalidated students are 6 times (635) as high as the sample weights for validated students (105.8). Therefore, estimates of overall aggregates, averages, and proportions have to be weighted estimates.

ESTIMATION OF SAMPLING ERROR

Student Characteristics

Every estimate of a total, $\hat{X}$, can be expressed as the sum of four components, as follows:

$$\hat{X} = \hat{X}_1 + \hat{X}_2 + \hat{X}_3 + \hat{X}_4 \qquad (1)$$

where $\hat{X}_1$ is the estimate from the certainty schools, $\hat{X}_2$ is the estimate from the ADS schools, $\hat{X}_3$ is the estimate from the RDS schools in the 25 largest SMSA's, and $\hat{X}_4$ is the estimate from the remaining schools. Since the samples were drawn independently in each stratum, the variance of $\hat{X}$ is the sum of the variances of the four components. That is,

$$\text{Var } \hat{X} = \text{Var } \hat{X}_1 + \text{Var } \hat{X}_2 + \text{Var } \hat{X}_3 + \text{Var } \hat{X}_4. \qquad (2)$$

The variance of a ratio $\hat{R} = \hat{Y}/\hat{X}$ may be approximated as

$$\text{Var } R \doteq \left[\frac{\hat{Y}}{\hat{X}}\right]^2 \left[\frac{\text{Var } \hat{X}}{\hat{X}^2} + \frac{\text{Var } \hat{Y}}{\hat{Y}^2} - \frac{2\text{Cov}(\hat{X},\hat{Y})}{\hat{X}\hat{Y}}\right].$$

where

$$\text{Cov}(\hat{X},\hat{Y}) = \text{Cov}(\hat{X}_1,\hat{Y}_1) + \text{Cov}(\hat{X}_2,\hat{Y}_2) + \text{Cov}(\hat{X}_3,\hat{Y}_3) + \text{Cov}(\hat{X}_4,\hat{Y}_4).$$

We examine these components separately.

In the certainty schools, the selection of students provides the only source of sampling error in our estimates. Within each school students were separated into two lists, validated and unvalidated, sorted by grant amount, and then each list was systematically sampled. An estimate of Var $\hat{X}_1$ is then

$$\text{Var } \hat{X}_1 = \sum_s \sum_{v \in s} \frac{n_v}{2(n_v - 1)} \sum_{i=1}^{n_v - 1} (w_i x_i - w_{i+1} x_{i+1})^2$$

where

- i denotes the individual case in its order of selection;
- w_i denotes the case weight;
- v specifies whether the case is validated or unvalidated;
- n_v is the number of validated (or unvalidated) students responding in a particular school; and
- s denotes the school.

Similarly, $\text{Cov}(\hat{X}_1, \hat{Y}_1)$ can be computed as

$$\sum_s \sum_{v \in s} \frac{n_v}{2(n_v - 1)} \sum_{i=1}^{n_v - 1} (w_i x_i - w_{i+1} x_{i+1})(w_i y_i - w_{i+1} y_{i+1})$$

Within a few schools, validated and unvalidated respondents were pooled to facilitate computation.

The sample of ADS schools was selected from a sorted list using systematic sampling with equal probability.

An estimate of Var $\hat{X}_2$ is then:

$$\text{Var } \hat{X}_2 = \frac{n_s}{2(n_s - 1)} \sum_{s=1}^{n_s-1} (x'_s - x'_{s+1})^2$$

where

s indexes a school in its order of selection;

n_s equals the number of ADS schools responding;

x'_s equals $\sum_{i \in s} w_i x_i$, the total for school s

i denotes the individual case; and

w_i denotes the case weight.

$\text{Cov}(\hat{X}_2, \hat{Y}_2)$ is computed as

$$\frac{n_s}{2(n_s - 1)} \sum_{s=1}^{n_s-1} (x'_s - x'_{s+1})(y'_s - Y'_{s+1})$$

The RDS schools within the 25 largest SMSA's were ordered by SMSA, control type, school type, and number of eligibles preparatory to selecting a preliminary sample of schools systematically with probability proportional to the number of eligibles. Although the sample was drawn with a single random start, for the purpose of variance estimation we assume that we have independent strata defined by SMSA and estimate Var $\hat{X}_3$ as follows:

$$\text{Var } \hat{X}_3 = \sum_{SMSA} \frac{n_{SMSA}}{n_{SMSA}-1} \sum_{s \in SMSA} (x'_s - \bar{x}'_{SMSA})^2$$

where

s denotes a school;

n_{SMSA} equals the number of schools responding from a particular SMSA;

x'_s equals $\sum_{i \in s} w_i x_i$, the total for school s;

$\bar{x}'_{SMSA}$ equals $\frac{1}{n_{SMSA}} \sum_{s \in SMSA} x'_s$, the average school level total within an SMSA;

i denotes the individual case; and

w_i denotes the case weight.

Cov $(\hat{X}_3 \hat{Y}_3)$ is computed as

$$\sum_{SMSA} \frac{n_{SMSA}}{n_{SMSA}-1} \sum_{s \in SMSA} (x'_s - \bar{x}'_{SMSA})(y'_s - \bar{y}'_{SMSA}).$$

In two cases there was only a single school responding in an SMSA. Where this happened, we pooled data from the SMSA adjacent in our sorted list to facilitate the calculations.

The RDS schools outside the 25 largest SMSA's were ordered and grouped into clusters of roughly equal size. Clusters were grouped into 40 strata. The first stage of sample selection chose two clusters per stratum with equal probability. An expression for the variance estimate Var $\hat{X}_4$ is

$$\text{Var } \hat{X}_4 = \sum_{h=1}^{40} \frac{n_h}{n_h - 1} \sum_{c \in h} (x'_c - \bar{x}'_h)^2$$

where

c denotes the cluster;

h denotes the stratum;

n_h equals the number of clusters responding from stratum h;

x'_c equals $\sum_{i \in c} w_i x_i$ the total for cluster c;

$\bar{x}'_h$ equals $\frac{1}{n_h} \sum_{c \in h} x'_c$ the average cluster total in stratum h;

i denotes the individual case; and

w_i denotes the case weight.

Similarly, $\text{Cov}(\hat{X}_4, \hat{Y}_4)$ is computed as

$$\sum_{h=1}^{40} \frac{n_h}{n_h - 1} \sum_{c \in h} (x'_c - \bar{x}'_h)(y'_c - y'_h).$$

Estimated Sampling Errors

In this section we present estimated sampling errors developed using the methods described in the previous section. For each statistic we present the estimate itself, the standard error of the estimate, and the coefficient of variation (standard error of the estimate divided by the estimate). Estimates presented in these tables will not agree strictly with estimates presented in the text of Volume 1 for two reasons. First, the estimates presented here were developed using software procedures different from the software utilized for Volume 1 estimates. These differences in computer rounding and truncation can be expected to account for large portions of the small differences. Second, development of the estimates presented here is based on a slightly different methodology concerning assumptions about missing or incomplete data.

Figure 2-4 presents the standard errors of estimated aggregate payment errors. The coefficients of variation for student and case error are always less than 11 percent, and the coefficient for total absolute case error is only 6 percent. Estimated standard errors for number of cases with payment or eligibility errrors are presented in Figure 2-5. The standard errors for estimated numbers of payment errors are generally below 10 percent. Occurrence of eligibility erorrs, except AEP and FAT, is extremely rare. For the course length requirement we uncovered

only one error in our entire sample. Therefore, the coefficients of variation will appear to be high. However, the standard errors are low enough so that reasonably tight confidence intervals exist.

Standard errors for average payment errors are presented in Figure 1-6. These standard errors are generally $20 or less; thus, 95 percent confidence intervals would generally have a width of $80 of less.

The remaining tables provide sampling errors for the various estimates for the following categories:

Type of Control
Type of Institution
Grant Type
Validation Status

Figures 2-7 to 2-18 are for ther various aggregate payment errors; Figure 2-19 to 2-24 are for numbers of cases with payment errors; Figures 2-25 to 2-36 contain sampling errors for average payment errors; and finally, Figures 2-37 to 2-45 are for the estimated number of cases with eligibility errors.

Comparing estimated coefficients of variation between the earlier 1978-79 study and the current study reveals that for many of the overall estimates the current study almost always yielded lower coefficients of variation. However, the differences in survey instrumentation, field work procedures, grant levels, etc., between the two studies imply that any comparisons such as these must be done with extreme caution.

Statistic	Estimate ($ millions)	Standard Error of Estimate	Coefficient of Variation
OVERPAYMENTS			
Institution	264.05	30.06	0.11
Student	263.44	18.96	0.07
Case	511.73	37.00	0.07
UNDERPAYMENTS			
Institution	93.65	12.85	0.14
Student	46.75	4.75	0.10
Case	125.77	13.31	0.11
NET PAYMENT ERROR			
Institution	170.40	33.00	0.19
Student	216.69	17.09	0.08
Case	385.96	37.50	0.10
ABSOLUTE PAYMENT ERROR			
Institution	357.71	32.38	0.09
Student	310.19	21.73	0.07
Case	637.51	41.07	0.06

FIGURE 2-4

ESTIMATED SAMPLING ERRORS:
SELECTED AGGREGATE DOLLAR ESTIMATES

Statistic	Estimate (Thousands)	Standard Error of Estimate	Coefficient of Variation
Institution Overpayment	597.60	55.25	0.09
Institution Underpayment	373.27	43.36	0.12
Student Overpayment	665.43	41.50	0.06
Student Underpayment	203.08	17.47	0.09
Case Overpayments	1,141.84	66.85	0.06
Case Underpayments	491.04	45.12	0.09
ELIGIBILITY ERRORS			
Academic Progress	28.58	8.56	0.30
Enrollment Status	2.27	1.19	0.52
Course Length	.68	.68	1.00
Degree Program	1.36	.96	0.71
Possesses B.A.	4.08	1.69	0.41
Missing AEP	92.90	25.11	0.27
Missing FAT	102.28	15.04	0.15
Loan Default	.91	.70	0.77
Citizenship	1.36	.96	0.71

FIGURE 2-5

ESTIMATED SAMPLING ERRORS:
SELECTED ESTIMATES OF CASES WITH ERRORS

Statistic	Estimate (Dollars)	Standard Error of Estimate	Coefficient of Variation
AVERAGE OVERPAYMENT			
Institution	442	34	0.08
Student	396	17	0.04
Case	448	20	0.05
AVERAGE UNDERPAYMENT			
Institution	251	19	0.08
Student	230	13	0.06
Case	256	15	0.06
AVERAGE NET ERROR			
Institution	74	14	0.19
Student	93	7	0.08
Case	167	16	0.10
AVERAGE ABSOLUTE ERROR			
Institution	156	13	0.08
Student	133	8	0.06
Case	276	14	0.05

FIGURE 2-6

ESTIMATED SAMPLING ERRORS:
SELECTED AVERAGE DOLLAR ESTIMATES

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	169.93	27.40	0.16
Private	55.68	12.77	0.23
Proprietary	38.44	12.07	0.31
TYPE OF INSTITUTION			
Four-Year	166.54	23.98	0.14
Two-Year	72.67	17.42	0.24
Other	24.84	9.89	0.40
GRANT TYPE			
Independent	124.75	16.70	0.13
Dependent	139.30	17.37	0.12
VALIDATION STATUS			
Non-Validated	251.82	28.53	0.11
Validated	12.23	2.12	0.17
ALL CATEGORIES COMBINED	264.05	30.06	0.11

FIGURE 2-7

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES: INSTITUTION OVERPAYMENT

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	161.05	16.34	0.10
Private	71.03	11.71	0.16
Proprietary	31.35	8.08	0.26
TYPE OF INSTITUTION			
Four-Year	190.09	18.07	0.10
Two-Year	53.56	7.79	0.15
Other	19.79	4.94	0.25
GRANT TYPE			
Independent	99.94	11.86	0.12
Dependent	163.50	12.98	0.08
VALIDATION STATUS			
Non-Validated	252.13	18.70	0.07
Validated	11.31	1.15	0.10
ALL CATEGORIES COMBINED	263.44	18.96	0.07

FIGURE 2-8

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
STUDENT OVERPAYMENT

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	323.04	35.13	0.11
Private	124.18	20.46	0.16
Proprietary	64.51	16.19	0.25
TYPE OF INSTITUTION			
Four-Year	350.26	30.84	0.09
Two-Year	121.05	21.81	0.18
Other	40.42	12.70	0.31
GRANT TYPE			
Independent	217.69	22.00	0.10
Dependent	294.04	22.34	0.08
VALIDATION STATUS			
Non-Validated	488.61	35.70	0.07
Validated	23.12	2.41	0.10
ALL CATEGORIES COMBINED	511.73	37.00	0.07

FIGURE 2-9

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES: CASE OVERPAYMENT

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	61.02	11.71	0.19
Private	5.72	1.46	0.26
Proprietary	26.90	6.69	0.25
TYPE OF INSTITUTION			
Four-Year	32.46	8.26	0.25
Two-Year	40.33	8.84	0.22
Other	20.86	6.61	0.32
GRANT TYPE			
Independent	49.05	9.07	0.18
Dependent	44.60	7.10	0.16
VALIDATION STATUS			
Non-Validated	90.18	12.73	0.14
Validated	3.46	0.56	0.16
ALL CATEGORIES COMBINED	93.65	12.85	0.14

FIGURE 2-10

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
INSTITUTION UNDERPAYMENT

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	29.26	3.90	0.13
Private	13.42	2.67	0.20
Proprietary	4.06	2.17	0.53
TYPE OF INSTITUTION			
Four-Year	34.77	4.21	0.12
Two-Year	8.60	1.63	0.19
Other	3.37	2.13	0.63
GRANT TYPE			
Independent	9.17	1.93	0.21
Dependent	37.58	4.31	0.11
VALIDATION STATUS			
Non-Validated	43.94	4.60	0.10
Validated	2.81	0.40	0.14
ALL CATEGORIES COMBINED	46.75	4.75	0.10

FIGURE 2-11

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES: STUDENT UNDERPAYMENT

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	82.95	12.55	0.15
Private	17.19	3.15	0.18
Proprietary	25.63	6.27	0.24
TYPE OF INSTITUTION			
Four-Year	60.70	9.55	0.16
Two-Year	44.64	9.16	0.21
Other	20.43	6.18	0.30
GRANT TYPE			
Independent	52.07	9.09	0.17
Dependent	73.70	8.31	0.11
VALIDATION STATUS			
Non-Validated	119.89	13.14	0.11
Validated	5.88	0.65	0.11
ALL CATEGORIES COMBINED	125.77	13.31	0.11

FIGURE 2-12

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASE OVERPAYMENT

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	108.91	28.44	0.26
Private	49.95	12.61	0.25
Proprietary	11.54	13.33	1.16
TYPE OF INSTITUTION			
Four-Year	134.08	25.39	0.19
Two-Year	32.34	17.42	0.54
Other	3.99	11.48	2.88
GRANT TYPE			
Independent	75.70	18.90	0.25
Dependent	94.70	19.18	0.20
VALIDATION STATUS			
Non-Validated	161.63	31.52	0.19
Validated	8.77	2.24	0.26
ALL CATEGORIES COMBINED	170.40	33.00	0.19

FIGURE 2-13

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NET INSTITUTION ERROR

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	230.96	31.09	0.13
Private	61.40	13.08	0.21
Proprietary	65.34	14.26	0.22
TYPE OF INSTITUTION			
Four-Year	199.01	25.33	0.13
Two-Year	113.00	21.43	0.19
Other	45.70	12.29	0.27
GRANT TYPE			
Independent	173.80	19.10	0.11
Dependent	183.91	18.33	0.10
VALIDATION STATUS			
Non-Validated	342.01	30.97	0.09
Validated	15.70	2.14	0.14
ALL CATEGORIES COMBINED	357.71	32.38	0.09

FIGURE 2-14

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
ABSOLUTE INSTITUTION ERROR

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	131.79	14.26	0.11
Private	57.61	10.54	0.18
Proprietary	27.29	7.50	0.27
TYPE OF INSTITUTION			
Four-Year	155.32	16.06	0.10
Two-Year	44.95	7.09	0.16
Other	16.42	4.27	0.26
GRANT TYPE			
Independent	90.77	12.18	0.13
Dependent	125.92	11.32	0.09
VALIDATION STATUS			
Non-Validated	208.19	16.99	0.08
Validated	8.50	1.07	0.13
ALL CATEGORIES COMBINED	216.69	17.09	0.08

FIGURE 2-15

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NET STUDENT ERROR

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	190.31	19.01	0.10
Private	84.46	13.32	0.16
Proprietary	35.42	9.15	0.26
TYPE OF INSTITUTION			
Four-Year	224.86	20.76	0.09
Two-Year	62.16	8.74	0.14
Other	23.17	6.31	0.27
GRANT TYPE			
Independent	109.12	11.84	0.11
Dependent	201.07	15.68	0.08
VALIDATION STATUS			
Non-Validated	296.07	21.29	0.07
Validated	14.12	1.35	0.10
ALL CATEGORIES COMBINED	310.19	21.73	0.07

FIGURE 2-16

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
ABSOLUTE STUDENT ERROR

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	240.09	33.33	0.14
Private	106.99	19.34	0.18
Proprietary	38.88	15.13	0.39
TYPE OF INSTITUTION			
Four-Year	289.56	29.49	0.10
Two-Year	76.41	20.61	0.27
Other	19.99	11.83	0.59
GRANT TYPE			
Independent	165.62	23.96	0.14
Dependent	220.33	22.24	0.10
VALIDATION STATUS			
Non-Validated	368.71	36.36	0.10
Validated	17.24	2.24	0.14
ALL CATEGORIES COMBINED	385.96	37.50	0.10

FIGURE 2-17

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
NET CASE ERROR

Category	Estimate ($ millions)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	405.99	40.89	0.10
Private	141.37	21.98	0.16
Proprietary	90.15	19.34	0.21
TYPE OF INSTITUTION			
Four-Year	410.96	34.86	0.08
Two-Year	165.69	26.36	0.16
Other	60.86	16.09	0.26
GRANT TYPE			
Independent	269.77	23.64	0.09
Dependent	367.75	25.32	0.07
VALIDATION STATUS			
Non-Validated	608.51	39.66	0.07
Validated	29.01	2.55	0.09
ALL CATEGORIES COMBINED	637.51	41.07	0.06

FIGURE 2-18

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
ABSOLUTE CASE ERROR

Category	Estimate (Thousands)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	465.91	56.58	0.12
Private	72.15	15.32	0.21
Proprietary	59.54	13.69	0.23
TYPE OF INSTITUTION			
Four-Year	338.31	36.19	0.11
Two-Year	216.30	46.19	0.21
Other	42.99	11.56	0.27
GRANT TYPE			
Independent	285.26	36.19	0.13
Dependent	312.34	26.58	0.09
VALIDATION STATUS			
Non-Validated	569.01	52.87	0.09
Validated	28.59	3.54	0.12
ALL CATEGORIES COMBINED	597.60	55.25	0.09

FIGURE 2-19

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH INSTITUTION OVERPAYMENT

Category	Estimate (Thousands)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	299.70	43.28	0.14
Private	17.13	3.87	0.23
Proprietary	56.44	11.42	0.20
TYPE OF INSTITUTION			
Four-Year	132.00	23.75	0.18
Two-Year	200.15	38.78	0.19
Other	41.11	10.22	0.25
GRANT TYPE			
Independent	185.99	28.38	0.15
Dependent	187.28	23.55	0.13
VALIDATION STATUS			
Non-Validated	357.28	42.50	0.12
Validated	15.99	1.69	0.11
ALL CATEGORIES COMBINED	373.27	43.36	0.12

FIGURE 2-20

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH INSTITUTION UNDERPAYMENT

Category	Estimate (Thousands)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	421.71	40.88	0.10
Private	182.25	21.72	0.12
Proprietary	61.46	14.00	0.23
TYPE OF INSTITUTION			
Four-Year	479.70	39.65	0.08
Two-Year	142.27	19.56	0.14
Other	43.47	10.74	0.25
GRANT TYPE			
Independent	158.53	14.03	0.09
Dependent	506.90	35.87	0.07
VALIDATION STATUS			
Non-Validated	635.61	40.79	0.06
Validated	29.82	2.54	0.09
ALL CATEGORIES COMBINED	665.43	41.50	0.06

FIGURE 2-21

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH STUDENT OVERPAYMENT

Category	Estimate (Thousands)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	122.17	14.52	0.12
Private	61.42	10.80	0.18
Proprietary	19.49	8.50	0.44
TYPE OF INSTITUTION			
Four-Year	150.28	14.51	0.10
Two-Year	36.82	7.37	0.20
Other	15.98	8.29	0.52
GRANT TYPE			
Independent	29.60	4.95	0.17
Dependent	173.47	16.12	0.09
VALIDATION STATUS			
Non-Validated	191.14	16.81	0.09
Validated	11.94	1.57	0.13
ALL CATEGORIES COMBINED	203.08	17.47	0.09

FIGURE 2-22

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH STUDENT UNDERPAYMENT

Category	Estimate (Thousands)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	803.14	72.79	0.09
Private	239.86	28.59	0.12
Proprietary	98.83	21.19	0.21
TYPE OF INSTITUTION			
Four-Year	753.20	52.76	0.07
Two-Year	320.83	52.33	0.16
Other	67.80	17.79	0.26
GRANT TYPE			
Independent	401.22	39.56	0.10
Dependent	740.62	43.21	0.06
VALIDATION STATUS			
Non-Validated	1,089.30	64.59	0.06
Validated	52.54	4.32	0.08
ALL CATEGORIES COMBINED	1,141.84	66.85	0.06

FIGURE 2-23

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH CASE OVERPAYMENT

Category	Estimate (Thousands)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	367.26	45.39	0.12
Private	71.18	11.21	0.16
Proprietary	52.60	11.86	0.23
TYPE OF INSTITUTION			
Four-Year	242.55	27.71	0.11
Two-Year	207.10	39.29	0.19
Other	41.38	11.41	0.28
GRANT TYPE			
Independent	183.66	28.27	0.15
Dependent	307.39	26.98	0.09
VALIDATION STATUS			
Non-Validated	466.26	44.09	0.09
Validated	24.78	2.28	0.09
ALL CATEGORIES COMBINED	491.04	45.12	0.09

FIGURE 2-24

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH CASE UNDERPAYMENT

Category	Estimate Dollars	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	365	36	0.10
Private	772	73	0.09
Proprietary	646	91	0.14
TYPE OF INSTITUTION			
Four-Year	492	41	0.08
Two-Year	336	49	0.15
Other	578	119	0.21
GRANT TYPE			
Independent	437	40	0.09
Dependent	446	37	0.08
VALIDATION STATUS			
Non-Validated	443	34	0.08
Validated	428	39	0.11
ALL CATEGORIES COMBINED	442	34	0.08

FIGURE 2-25

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE INSTITUTION OVERPAYMENT.

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	382	17	0.04
Private	390	34	0.09
Proprietary	510	61	0.12
TYPE OF INSTITUTION			
Four-Year	396	20	0.05
Two-Year	376	29	0.08
Other	455	54	0.12
GRANT TYPE			
Independent	630	44	0.07
Dependent	323	13	0.04
VALIDATION STATUS			
Non-Validated	397	17	0.04
Validated	379	24	0.06
ALL CATEGORIES COMBINED	396	17	0.04

FIGURE 2-26

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE STUDENT OVERPAYMENT

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	402	22	0.06
Private	518	40	0.08
Proprietary	653	51	0.08
TYPE OF INSTITUTION			
Four-Year	465	24	0.05
Two-Year	377	35	0.09
Other	596	52	0.09
GRANT TYPE			
Independent	543	37	0.07
Dependent	397	19	0.05
VALIDATION STATUS			
Non-Validated	449	21	0.05
Validated	440	29	0.07
ALL CATEGORIES COMBINED	448	20	0.05

FIGURE 2-27

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES: AVERAGE CASE OVERPAYMENT

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	204	19	0.09
Private	334	59	0.18
Proprietary	477	60	0.13
TYPE OF INSTITUTION			
Four-Year	246	35	0.14
Two-Year	201	17	0.09
Other	507	67	0.13
GRANT TYPE			
Independent	264	24	0.09
Dependent	238	24	0.10
VALIDATION STATUS			
Non-Validated	252	20	0.08
Validated	217	28	0.13
ALL CATEGORIES COMBINED	251	19	0.08

FIGURE 2-28

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE INSTITUTION UNDERPAYMENT

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	240	20	0.08
Private	219	24	0.11
Proprietary	208	31	0.15
TYPE OF INSTITUTION			
Four-Year	231	18	0.08
Two-Year	234	24	0.10
Other	211	35	0.17
GRANT TYPE			
Independent	310	46	0.15
Dependent	217	15	0.07
VALIDATION STATUS			
Non-Validated	230	14	0.06
Validated	236	21	0.09
ALL CATEGORIES COMBINED	230	13	0.06

FIGURE 2-29

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE STUDENT UNDERPAYMENT

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	226	16	0.07
Private	242	23	0.10
Proprietary	487	59	0.12
TYPE OF INSTITUTION			
Four-Year	250	20	0.08
Two-Year	216	17	0.08
Other	494	71	0.14
GRANT TYPE			
Independent	284	24	0.08
Dependent	240	16	0.07
VALIDATION STATUS			
Non-Validated	257	15	0.06
Validated	237	20	0.09
ALL CATEGORIES COMBINED	256	15	0.06

FIGURE 2-30

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE CASE UNDERPAYMENT

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	65	17	0.26
Private	117	25	0.22
Proprietary	61	65	1.07
TYPE OF INSTITUTION			
Four-Year	92	18	0.20
Two-Year	46	24	0.52
Other	30	83	2.75
GRANT TYPE			
Independent	86	21	0.25
Dependent	67	14	0.20
VALIDATION STATUS			
Non-Validated	74	14	0.20
Validated	81	20	0.24
ALL CATEGORIES COMBINED	74	14	0.19

FIGURE 2-31

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE NET INSTITUTION ERROR

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	138	14	0.10
Private	144	25	0.18
Proprietary	344	42	0.12
TYPE OF INSTITUTION			
Four-Year	136	16	0.12
Two-Year	160	20	0.12
Other	347	39	0.11
GRANT TYPE			
Independent	197	17	0.08
Dependent	130	13	0.10
VALIDATION STATUS			
Non-Validated	156	13	0.08
Validated	145	18	0.12
ALL CATEGORIES COMBINED	156	13	0.08

FIGURE 2-32

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE ABSOLUTE INSTITUTION ERROR

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	78	7	0.09
Private	132	16	0.12
Proprietary	139	33	0.24
TYPE OF INSTITUTION			
Four-Year	106	9	0.09
Two-Year	63	9	0.15
Other	115	30	0.26
GRANT TYPE			
Independent	101	14	0.14
Dependent	88	6	0.07
VALIDATION STATUS			
Non-Validated	94	7	0.08
Validated	78	9	0.11
ALL CATEGORIES COMBINED	93	7	0.08

FIGURE 2-33

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE NET STUDENT ERROR

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	113	8	0.07
Private	194	17	0.09
Proprietary	180	28	0.16
TYPE OF INSTITUTION			
Four-Year	153	10	0.07
Two-Year	87	10	0.12
Other	162	21	0.13
GRANT TYPE			
Independent	122	14	0.12
Dependent	141	7	0.05
VALIDATION STATUS			
Non-Validated	134	8	0.06
Validated	129	10	0.08
ALL CATEGORIES COMBINED	133	8	0.06

FIGURE 2-34

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE ABSOLUTE STUDENT ERROR

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	142	18	0.12
Private	250	30	0.12
Proprietary	204	57	0.28
TYPE OF INSTITUTION			
Four-Year	198	20	0.10
Two-Year	107	27	0.25
Other	151	68	0.45
GRANT TYPE			
Independent	186	27	0.15
Dependent	155	15	0.09
VALIDATION STATUS			
Non-Validated	168	16	0.10
Validated	157	19	0.12
ALL CATEGORIES COMBINED	167	16	0.10

FIGURE 2-35

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES: AVERAGE NET CASE ERROR

Category	Estimate (Dollars)	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	240	15	0.06
Private	330	28	0.09
Proprietary	473	39	0.08
TYPE OF INSTITUTION			
Four-Year	281	18	0.07
Two-Year	233	22	0.09
Other	459	31	0.07
GRANT TYPE			
Independent	303	22	0.07
Dependent	259	13	0.05
VALIDATION STATUS			
Non-Validated	277	15	0.05
Validated	265	17	0.06
ALL CATEGORIES COMBINED	276	14	0.05

FIGURE 2-36

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
AVERAGE ABSOLUTE CASE ERROR

Category	Estimate	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	22,797	8,543	0.37
Private	2,265	1,189	0.52
Proprietary	3,513	1,803	0.51
TYPE OF INSTITUTION			
Four-Year	21,302	8,280	0.39
Two-Year	5,102	2,415	0.48
Other	2,153	1,525	0.71
GRANT TYPE			
Independent	12,714	5,333	0.42
Dependent	15,862	4,058	0.26
VALIDATION STATUS			
Non-Validated	27,203	8,357	0.31
Validated	1,372	462	0.34
ALL CATEGORIES COMBINED	28,575	8,562	0.30

FIGURE 2-37

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH ACADEMIC PROGRESS ELIGIBILITY ERRORS

Category	Estimate	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	2,153	1,183	0.55
Private	0	0	.
Proprietary	112	112	1.00
TYPE OF INSTITUTION			
Four-Year	1,360	962	0.71
Two-Year	793	689	0.87
Other	112	112	1.00
GRANT TYPE			
Independent	1,473	968	0.66
Dependent	793	689	0.87
VALIDATION STATUS			
Non-Validated	2,040	1,178	0.58
Validated	225	159	0.71
ALL CATEGORIES COMBINED	2,265	1,189	0.52

FIGURE 2-38

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH ENROLLMENT STATUS ELIGIBILITY ERRORS

Category	Estimate	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	680	680	1.00
Private	0	0	.
Proprietary	0	0	.
TYPE OF INSTITUTION			
Four-Year	680	680	1.00
Two-Year	0	0	.
Other	0	0	.
GRANT TYPE			
Independent	0	0	.
Dependent	680	680	1.00
VALIDATION STATUS			
Non-Validated	680	680	1.00
Validated	0	0	.
ALL CATEGORIES COMBINED	680	680	1.00

FIGURE 2-39

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH COURSE LENGTH ELIGIBILITY ERRORS

Category	Estimate	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	1,360	962	0.71
Private	0	0	.
Proprietary	0	0	.
TYPE OF INSTITUTION			
Four-Year	680	680	1.00
Two-Year	680	680	1.00
Other	0	0	.
GRANT TYPE			
Independent	0	0	.
Dependent	1,360	962	0.71
VALIDATION STATUS			
Non-Validated	1,360	962	0.71
Validated	0	0	.
ALL CATEGORIES COMBINED	1,360	962	0.71

FIGURE 2-40

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH DEGREE STUDENT ELIGIBILITY ERRORS

Category	Estimate	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	2,720	1,388	0.51
Private	1,360	962	0.71
Proprietary	0	0	.
TYPE OF INSTITUTION			
Four-Year	2,720	1,360	0.50
Two-Year	1,360	1,001	0.74
Other	0	0	.
GRANT TYPE			
Independent	2,040	1,210	0.59
Dependent	2,040	1,178	0.58
VALIDATION STATUS			
Non-Validated	4,081	1,689	0.41
Validated	0	0	.
ALL CATEGORIES COMBINED	4,081	1,689	0.41

FIGURE 2-41

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH B.A. ELIGIBILITY ERRORS

Category	Estimate	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	51,453	21,947	0.43
Private	22,911	7,736	0.34
Proprietary	18,535	8,017	0.43
TYPE OF INSTITUTION			
Four-Year	61,903	23,231	0.38
Two-Year	16,542	6,177	0.37
Other	14,454	6,455	0.45
GRANT TYPE			
Independent	35,691	12,933	0.36
Dependent	57,207	15,008	0.26
VALIDATION STATUS			
Non-Validated	87,951	23,697	0.27
Validated	4,948	1,783	0.36
ALL CATEGORIES COMBINED	92,899	25,114	0.27

FIGURE 2-42

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH A.E.P. MISSING

Category	Estimate	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	66,449	15,260	0.23
Private	24,951	6,552	0.26
Proprietary	10,876	4,380	0.40
TYPE OF INSTITUTION			
Four-Year	63,386	12,453	0.20
Two-Year	29,374	7,628	0.26
Other	9,516	4,209	0.44
GRANT TYPE			
Independent	57,637	10,204	0.18
Dependent	44,639	7,690	0.17
VALIDATION STATUS			
Non-Validated	98,340	14,653	0.15
Validated	3,936	854	0.22
ALL CATEGORIES COMBINED	102,276	15,042	0.15

FIGURE 2-43

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES MISSING F.A.T.

Category	Estimate	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	905	698	0.77
Private	0	0	.
Proprietary	0	0	.
TYPE OF INSTITUTION			
Four-Year	793	689	0.87
Two-Year	112	112	1.00
Other	0	0	.
GRANT TYPE			
Independent	680	680	1.00
Dependent	225	159	0.71
VALIDATION STATUS			
Non-Validated	680	680	1.00
Validated	225	159	0.71
ALL CATEGORIES COMBINED	905	698	0.77

FIGURE 2-44

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH LOAN DEFAULT

Category	Estimate	Standard Error	Coefficient of Variation
TYPE OF CONTROL			
Public	1,360	962	0.71
Private	0	0	.
Proprietary	0	0	.
TYPE OF INSTITUTION			
Four-Year	1,360	962	0.71
Two-Year	0	0	.
Other	0	0	.
GRANT TYPE			
Independent	680	680	1.00
Dependent	680	680	1.00
VALIDATION STATUS			
Non-Validated	1,360	962	0.71
Validated	0	0	.
ALL CATEGORIES COMBINED	1,360	962	0.71

FIGURE 2-45

ESTIMATED SAMPLING ERRORS BY SELECTED CATEGORIES:
CASES WITH CITIZENSHIP ELIGIBILITY ERRORS

SURVEY RESPONSE RATES AND POTENTIAL NONRESPONSE BIAS

In the following section we discuss response rates for the survey of parents and students and review the existing data for potential nonresponse bias. As shown in the following section, the response rate was high and any problem with nonresponse bias is likely to be of minor significance.

Response Rates for the Student/Parent BEOG Survey

The fieldwork carried out by Westat resulted in completed questionnaires sufficient to verify all application form entries for 90 percent of the sampled grants. Details concerning this response rate, as well as an alternative calculation, are included below.

Defining response rates for the BEOGs Quality Control study presents a unique problem because a case or grant could involve two separate questionnaires. For a dependent student, it is necessary to have a completed questionnaire from both the student and parent in order to verify all the application entries. Furthermore, since both signed the application, their cooperation is mandatory.

For independent student grants, it is only necessary to have a completed student questionnaire to completely verify the application entries. Those entries concerning parental support which are used to establish independent status are verifiable using the student questionnaire responses.

The response rates shown in Figure 2-46 are based on three general disposition categories--completed; unavailable for interview; and refusers, avoiders, breakoffs, etc. Figure 2-47 lists which detailed disposition codes have been combined in forming these three general categories.

The rates are all based on the number of original sampling points and questionnaire dispositions presented in Figure 2-47. The 3 response rates for dependent student grants are all above 90 percent; nearly 95 percent for student questionnaires; and almost 94 percent for parents. For 91 percent of the dependent student grants we were able to get completed interviews with both the parent and student.

Response rates for independent student grants are not as high, but for students we were able to complete over 87 percent of the interviews. Since the parent interview was voluntary for independent student grants, the degree of cooperation was expected to be low. These expectations were borne out as indicated by the 61 percent response rate for independent parent questionnaires. Of course, the percent of cases for which we completed a student and parent questionnaire also reflects this low level of cooperation, as indicated by the 58 percent paired completion rate. As indicated by the footnotes in Figure 2-46, many parents were deceased, out of the country, or could not be located.

The final two response rates are based on two methods of combining dependent and independent grants. The first entry is

Grant type and form of questionnaires	Completed	Nonrespondents	
		Unavailable for Interview	Refusers, Avoiders, Breakoffs, etc.
Dependent Student Grants:			
Student Questionnaire	94.6	1.0	4.4
Parent Questionnaire	93.6	1.9	4.4
Student and Parent	91.4	2.2	6.4*
Independent Student Grants:			
Student Questionnaire	87.4	1.0	11.6
Parent Questionnaire	61.0	15.6**	23.4***
Student and Parent	58.4	15.3*	26.3
Dependent and Independent Grants:			
Student and Parent	79.5	6.9	13.6*
Student and Parent for Dependent Student Grants and Student Questionnaire for Independent Student Grants	90.0	1.8	8.3*

FIGURE 2-46

RESPONSE ~~RATES~~

*A grant is included in this category if parent, student, or both were refusers, avoiders, breakoffs, etc.

**Nearly 52 percent (137 cases) of this 15.6 percent consist of deceased parents; 36 percent (96 cases) consist of parents who are out of the country.

***Over 41 percent (162 cases) of this 23.4 percent consist of parents who could not be located.

A. Dependent Student Grants

Parent Questionnaire Disposition	Student Questionnaire Disposition			
	Completes	Nonrespondents		Totals
		Unavailable for Interview	Refusers, Avoiders, Break-offs, etc.	
Completes	2,732	15	53	2,800
Nonrespondents: Unavailable	40	12	5	57
Nonrespondents: Refusers, Avoiders, Breakoffs, etc.	57	2	74	133
Totals	2,829	29	132	2,990

B. Independent Student Grants

Parent Questionnaire Disposition	Student Questionnaire Disposition			
	Completes	Nonrespondents		Totals
		Unavailable for Interview	Refusers, Avoiders, Break-offs, etc.	
Completes	985	4	40	1,029
Nonrespondents: Unavailable	245	9	10	264
Nonrespondents: Refusers, Avoiders, Breakoffs, etc.	245	4	145	394
Totals	1,475	17	195	1,687

FIGURE 2-47

PARENT QUESTIONNAIRE DISPOSITIONS BY

General Category	Detailed Disposition Categories
Completes	11 - Interview Completed
Unavailable for Interview	14 - Extended illness 20 - Sampling error 21 - Out of country 22 - Deceased 35 - Away for field period
Refusers, Avoiders, Breakoffs, etc.	12 - Maximum Calls 13 - Cannot Locate: Address from School 15 - Refusal/Breakoff 16 - Avoider 17 - Language Problem 18 - Other 31 - Cannot Locate: No address from School 32 - Other member refused 33 - Refused because student quit 34 - No involvement with student, independent parents only.

FIGURE 2-48

DISPOSITION CATEGORIES

based on the demanding paired completion for both grant types. Its value of about 80 percent reflects the problem of getting independent parents to cooperate.

The last entry of 90 percent is most reflective of the study requirements. Here we used the paired completes for dependent student grants and student completes for independent student grants. Thus, for 90 percent of the grants we have questionnaire response sufficient to verify all application form entries.

Analysis of Nonresponse Bias

The purpose of this analysis of nonresponse in the survey of BEOG grant recipients is to determine the possible nonresponse bias in estimating program-wide grant error amounts. Even though this survey, as noted in the previous section, achieved very high parent and student response rates (90 percent overall), a careful analysis of the possible impact of the missing nonrespondents is important to the credibility of the study's eventual findings and conclusions.

The fieldwork carried out by Westat resulted in completed questionnaires sufficient to verify all application form entries for 90 percent of the sampled grants. This composite rate can be broken down into separate rates for independent and dependent student grants as follows: Questionnaires were completed for both students and parents for 91.4 percent of the dependent student grants. At least a student questionnaire was completed for 87.4 percent of the independent student grants.

Defining response rates for the BEOG Quality Control study presents a unique problem because a case or grant could involve two separate questionnaires.[1] For a dependent student, it is necessary to have a completed questionnaire from both the student and parent in order to verify all the application entries. Furthermore, since both signed the application, their cooperation is mandatory. For independent student grants, it is only necessary to have a completed student questionnaire to completely verify the application entries. Those entries concerning parental support which are used to establish independent status are verifiable using the student questionnaire responses.

Even with response rates this high, there is the possibility that the respondents may be quite different from nonrespondents in terms of disbursement error. While error in disbursements is the focus of this Quality Control study, for nonrespondents it is not possible to calculate grant error amount because we lack the necessary verification data, which is normally obtained from interviewing respondents. However, we can test the hypothesis that "nonrespondents are students who misrepresented their financial circumstances in order to obtain a substantially larger grant disbursement amount." If this hypothesis were true, we would expect nonrespondents to have received higher disbursements

---

[1] In addition, data collected through the Student Record Abstracts [SRA] could provide data sufficient to calculate a grant error.

than respondents. This could be tested by comparing the average disbursement amount for respondents versus nonrespondents. However, data on actual disbursements were not available when this analysis was performed. Available data on fall 1980 values of the student eligibility index (SEI), scheduled Basic Grant, and expected disbursements are expected to be highly corrrelated with actual disbursement and will be used in this analysis as proxies for actual disbursements.

In later sections we compare the average values for these three proxy variables for respondents and nonrespondents to ascertain the extent to which the differences would be symptomatic of misrepresentation of financial circumstances for the purpose of obtaining a substantially higher disbursement. While minor differences do exist, their levels are not sufficiently high to seriously undermine subsequent analysis. Furthermore, while nonrespondent dependent student cases have higher expected disbursements than respondents, the opposite is true for independent student cases, e.g., the potential biases are in the opposite directions. Finally, the average differences are often within the tolerances established for the student validation procedures.[1] Thus, the evidence currently available does not

---

[1] The effective tolerance is about $50 in terms of expected disbursement. However, this tolerance is limited to only a subset of the application items which affect the grant level.

support the hypotheses that nonrespondents are receiving substantially larger grants than respondents.[1]

In this report we utilize data collected in the fall of 1980. Spring 1981 data on enrollment status, actual and planned disbursements, and more recently calculated scheduled entitlements had already been collected; however, this data set was not available for analysis. It is expected that an analysis using the more recent data would yield identical conclusions with respect to nonresponse bias.

Data Base Description

The data base for this analysis consists of elements from four basic data sources as shown below:

BASIC DATA SOURCE	ELEMENTS
Fall Student Eligibility Report [SER]	Student Eligibility Index [SEI] Scheduled Basic Grant Expected Disbursement Grant Status (Independent or Dependent)
Student Questionnaire	Student Questionnaire Disposition
Parent Questionnaire	Parent Questionnaire Disposition
Master Sampling File	Sampling Weight

Records from these four files were merged according to case identification numbers. The final merged file contains about 4,700 records, 1 for each originally sampled grant.

---

[1] Note that it is still possible that nonrespondents might have larger grant errors than respondents; but data are not available to make such a test.

The SEI, scheduled grant, and expected disbursements are taken from the SER submitted by schools last fall. In other parts of the main analysis report, the SEI taken from school files in the spring is analyzed. As a consequence, average scheduled grants and expected disbursements used in this analysis will be somewhat higher than their spring counterparts because students tend to reduce credit hours or drop out as the academic year proceeds.

The purpose of this analysis is to assess differences in SEI, grants, and expected disbursements between respondents and nonrespondents; however, these values are oftentimes missing for respondents and nonrespondents as indicated in Figure 2-49.

Figure 2-49 indicates that the chance of missing an SEI, grant, or expected disbursement is higher for nonrespondents than for respondents. Less than 1 percent of the respondent cases had missing SEI: nonrespondents had missing SEI in 1.8 percent of the cases. Missing expected disbursements occurred in 25 percent of the respondent cases and 40 percent of the nonrespondent cases. These differential missing data rates between respondents and nonrespondents somewhat limit the strength of any conclusions which can be drawn from the following analysis.

Discussion of Results

In this section the weighted average values for SEI, scheduled grant, and expected disbursement are presented for respondents and nonrespondents.

Presence of Data Element	Data Element					
	Student Eligibility Index		Scheduled Grant		Expected Disbursement	
	Respondents	Nonrespondents	Respondents	Nonrespondents	Respondents	Nonrespondents
Element Present	4166	447	3193	288	3144	275
Element Missing	37	8	1012	167	1059	180
Total Cases	4203	455	4205	455	4203	455
Cases with % Missing Data Element	0.9	1.8	24.1	36.7	25.2	39.6

FIGURE 2-49

INCIDENCE OF MISSING DATA ELEMENT VALUES

These measures do not represent the ultimate focus of this Quality Control study: measuring the levels of error in actual disbursements is its major purpose. We will not be able to measure errors in the above three magnitudes for nonrespondents; thus, we are left with assessing differences in their levels as a measure of potential nonresponse bias.

Figure 2-50 presents information concerning expected disbursements. The mean expected disbursement is $985.2 as indicated in the left-most box. The next two boxes present the average values for respondents and nonrespondents with the circle between the two boxes containing the difference between the respective means. Nonrespondents had expected disbursements about $28 higher than respondents. Independent student nonrespondents had expected disbursements lower by $32, while dependent student nonrespondents had expected disbursements higher by $67. None of these three differences is statistically different from zero. Furthermore, their levels are not large in a substantive sense. The $28 difference is safely within the tolerances used in the BEOG validation procedures.

Average scheduled grants are presented in Figure 2-51. The overall difference of $53 and the difference, $89, for dependent students are statistically different from zero; however, their levels are not much above the $50 tolerance used in the BEOG's validation procedures. The difference of $10 for independent students is of little consequence.

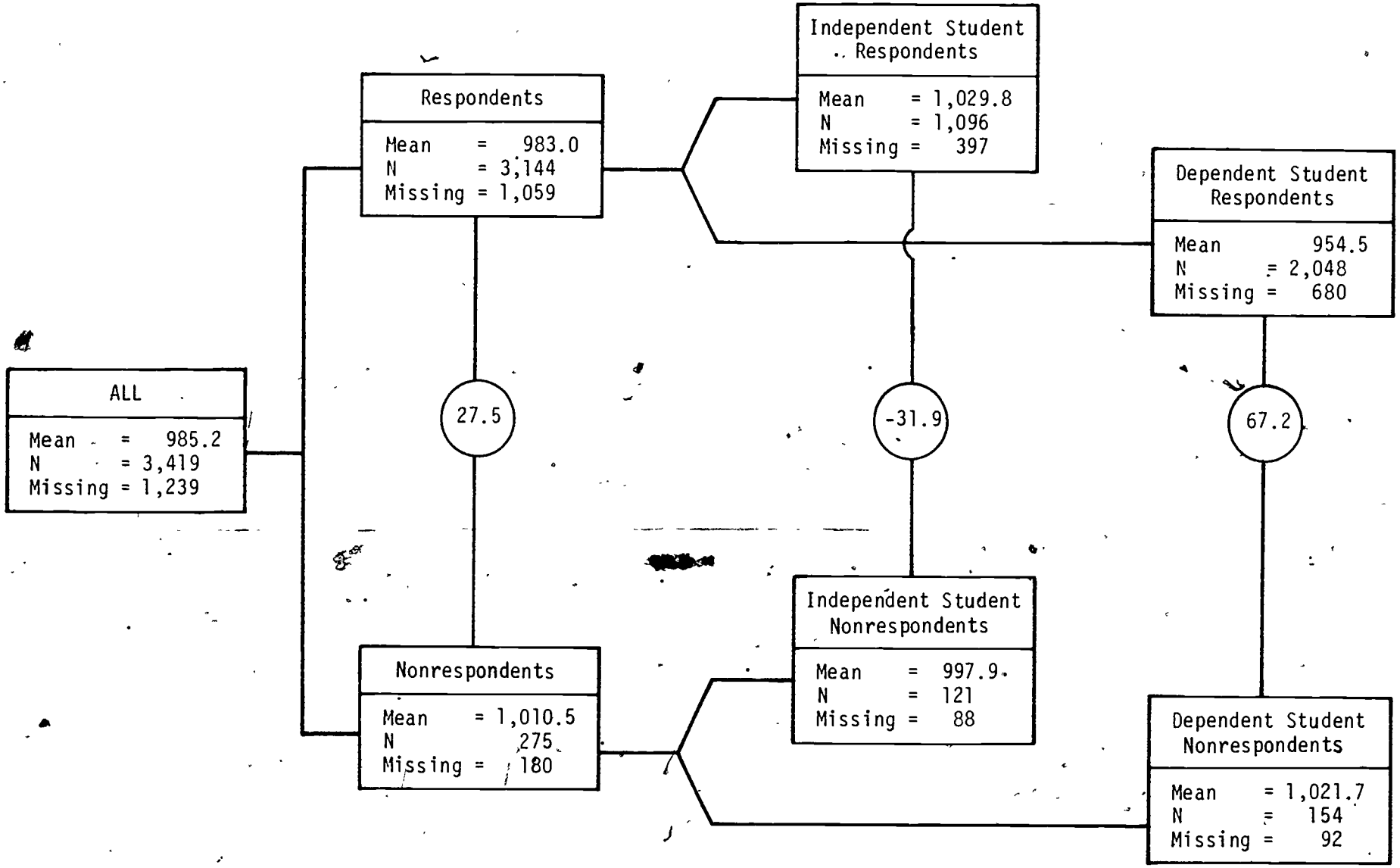


FIGURE 2-50

AVERAGE EXPECTED DISBURSEMENTS BY CASE DISPOSITION AND GRANT STATUS

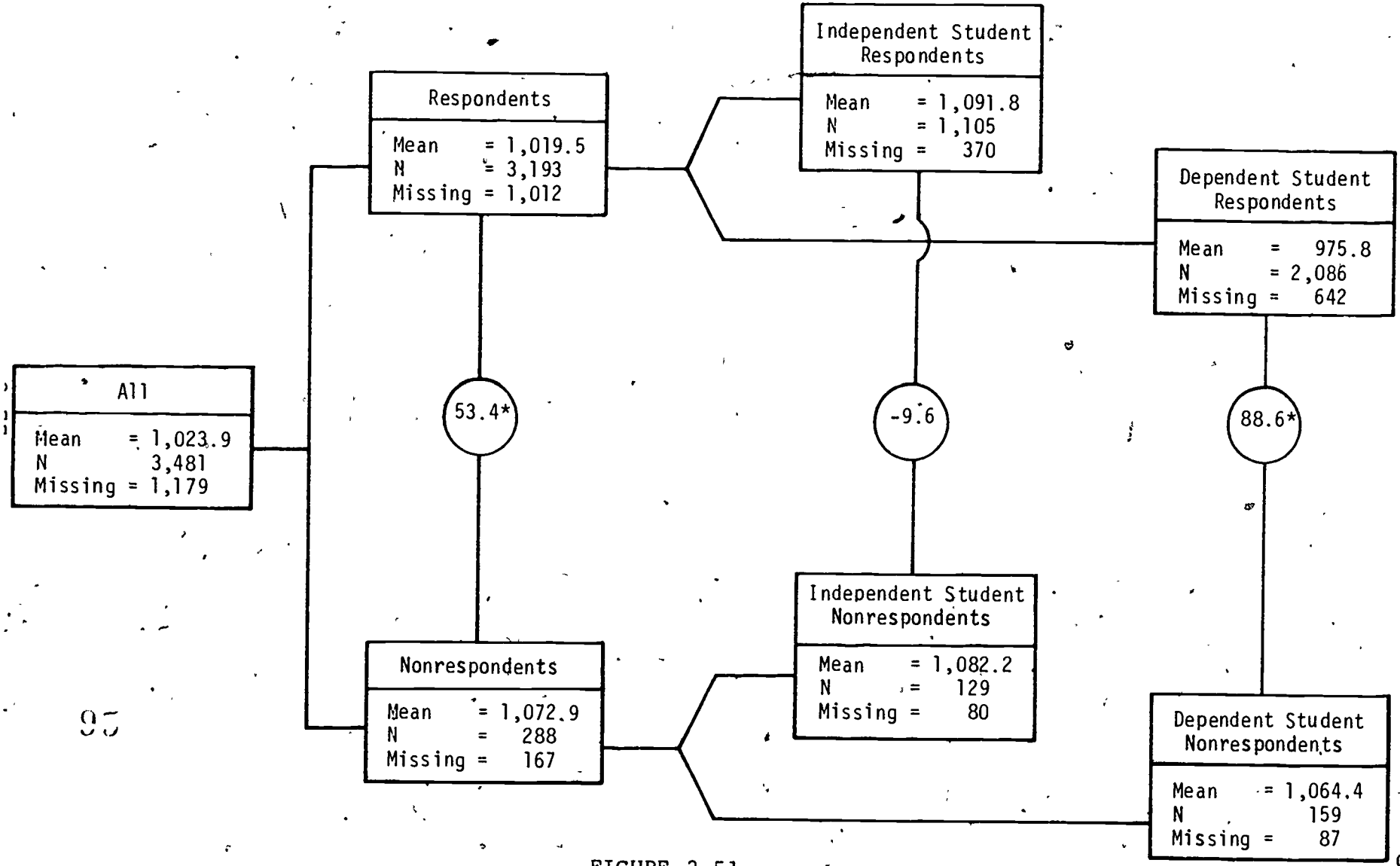


FIGURE 2-51

AVERAGE SCHEDULED GRANT BY CASE DISPOSITION AND GRANT STATUS

The greatest disparities or differences are for the average student eligibility index as presented in Figure 2-52. Here the difference for all grant types is 120 points, and for dependent students it is equally high. It is interesting that the effect of large differences in SEI is rather strongly mitigated by the grant determination procedures. Many grants and expected disbursements are not affected by changes in the SEI because of the three-part grant determination procedure.

Since the ultimate objective of the BEOG Quality Control study should be disbursement errors, it is reassuring that differences in average expected disbursements are so low. It is our opinion that the data base is not affected seriously enough by potential nonresponse bias to have any substantive impact on policy conclusions.

Further Analysis of Potential Nonresponse Bias

As we have already stated, we do not feel that the data base is seriously affected by nonresponse. To further strengthen this conclusion we have performed sensitivity analyses as presented in the following. If there is no nonresponse bias, as we contend, inferences drawn from respondents' data would also be expected to hold for nonrespondents. As a result, estimation of totals will be based on a straightforward adjustment using this assumption. However, there are alternative assumptions which can be made with regard to nonrespondents, as depicted in Figure 2-53.

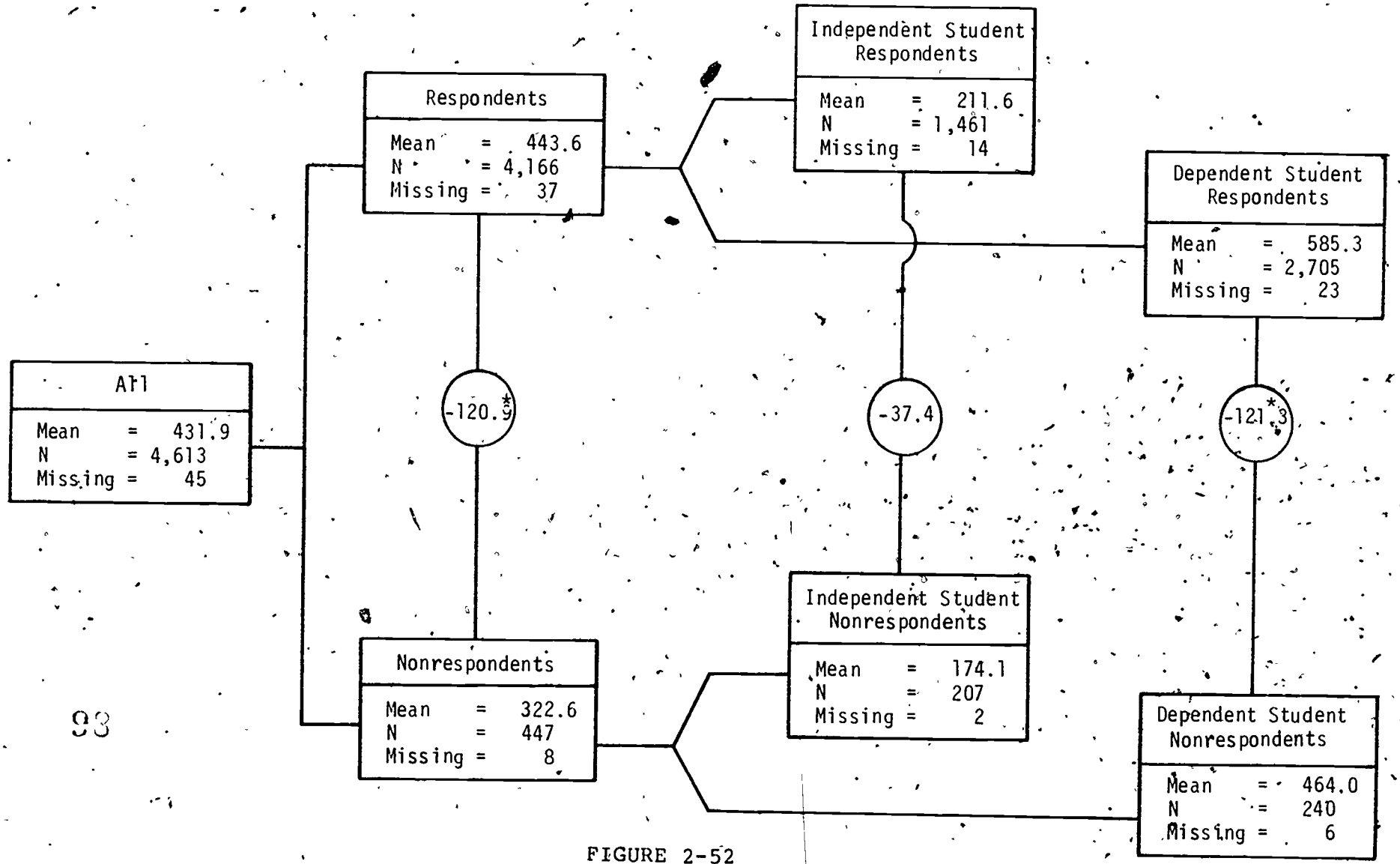


FIGURE 2-52

AVERAGE STUDENT ELIGIBILITY INDEX BY CASE DISPOSITION AND GRANT STATUS

SCENARIO	ESTIMATE OF TOTAL NET ERROR ($ MILLION)	AVERAGE ERROR ASSUMED FOR NONRESPONDENTS
Respondents and nonrespondents have equal net error	$453	$192
Entire value of expected disbursement is in error	563	1,011
Nonrespondents had errors equal to:		
95th percentile	578	1,126
90th percentile	540	838
75th percentile	471	326
25th percentile	427	0
10th percentile	402	-186
5th percentile	380	-350

2-81

FIGURE 2-53

SENSITIVITY OF TOTAL NET ERROR ESTIMATE USING ALTERNATIVE ASSUMPTIONS CONCERNING NONRESPONDENTS

The first entry of $453 million is based on the assumption that average net error is the same for respondents and nonrespondents. The second largest error estimate of $563 million is based on the assumption that the entire expected disbursement for nonrespondents is in error.

The largest estimate based on the 95th percentile, $578 million, assumes that the average error for nonrespondents, $1,126, exceeds their average expected disbursement--a somewhat unrealistic assumption.

Overall, the estimate used in the main report, $453 million, could be underestimated by $110 million only if the average error for nonrespondents equaled their entire expected disbursement.

The problem to be avoided is concluding that there is no error when indeed there is error, e.g., downward bias. In our case, the baseline estimate is high enough that even if it is biased downward by $100 million, the magnitude of the estimated error justified the conclusion that there is significant error in the Basic Grant program.

CHAPTER 3

DATA COLLECTION AND QUALITY CONTROL

Data were collected during the spring of 1981 from a nationally representative sample of 305 institutions and from approximately 4,500 BEOG recipients and their parents. In all, eight data sets, displayed in Figure 3-1 below, were collected. This chapter is organized into three sections. First, the student and parent survey, conducted by Westat's nationwide network of interviewers, is described. Second, the collection of secondary data used to verify student application data is discussed. In the final section, the institutional survey conducted by Advanced Technology is described.

Data Set	Data Source
1. Student File	Student Questionnaire: obtained from in-field interviews with sampled students
2. Parent File	Parent Questionnaire: obtained from in-field interviews with parents of sampled students
3. SER File	Student Eligibility Reports obtained from sampled institutions
4. IRS File	Certified 1979 IRS 1040 and 1040A forms, obtained through releases from sampled students and parents
5. Tax Assessor Records	Obtained through releases from sampled students and parents
6. Financial Institution Records	Obtained through releases from sampled students and parents
7. Institution File	Institutional Questionnaire: obtained from interviews with financial aid administrators at sampled institutions
8. Student Record Abstract File	Student Record Abstract: obtained from student file reviews at sampled insititutions

FIGURE 3-1

BEOG QUALITY CONTROL DATA SETS AND SOURCES

STUDENT/PARENT INTERVIEWS

The following aspects of the field data collection which will be discussed below include:

- Questionnaire development and pretest
- Field management structure
- Interviewer recruitment
- Training
- Field operations issues
- Reporting and the Automated Survey Control System
- Field problems

Questionnaire Development and Pretest

The basis for the first draft of the student/parent data collection instrument was the 1978-79 questionnaire revised by recommendations from Department of Education (ED), Advanced Technology, and Westat project staff. Eight drafts of the questionnaire were subsequently produced before the final questionnaires were printed. Pretests of draft II and draft IV provided useful information on the effectiveness of the questionnaires.

The pretest of draft II took place in the week of October 22-24. A brief training session was held which included the following topics:

- Westat questionnaire format conventions and general interviewing techniques
- Question-by-question specifications for the newly revised questionnaire
- Documentation requirements

Most of the Advanced Technology and Westat BEOG project staff interviewed in the pretest along with a professional pretest interviewer from Westat. Pretest respondents were selected from Montgomery County College and Howard University. They and their parents were contacted by telephone by a Westat interviewer who explained the study, screened the respondent for dependency status, and scheduled an appointment. Nine independent students, nine dependent students, nine dependent parents, and one independent parent was interviewed.

A debriefing session was held at the end of the week for pretest participants. The following topics were covered:

- Questionnaire administration time
- Flow of interview
- Ease of questionnaire use including instructions and format
- Clarity of questions, clarity of responses, and coding, question by question
- Other issues which participants wished to discuss

It was very clear from the pretest that: (1) the instrument needed to be divided into parent and student questionnaires; (2) the format for Section B (verification of application form items) needed to be simplified; and (3) the instructions and phrasing of questionnaires needed to be simplified.

Draft IV was produced from the results of the pretesting of Draft II and from comments by ED, Advanced Technology, and Westat project staff. Important revisions first evident in this draft were: (1) the emphasis of the time reference in each question and

(2) the creation of a "problem question" as the first question in each instrument. Draft IV was pretested by the pretest interviewer (4 interviews). The results indicated that the instrument flowed very well. The revisions which were made to subsequent drafts were relatively minor. (If more detail about the process is desired, see Westat Methodological Report.)

Suggestions for future BEOG Quality Control questionnaires based on our field experience include:

- Obtain more information about the role of the financial aid officer in completing application forms. Apparently there are instances in which he or she filled out the application for a student.
- Limit the questions asked of the parents of independent students. Those who were not involved in filling out the form became annoyed or thought it was a waste of money to conduct an in-person interview for so few questions.

Field Management Structure

To efficiently manage the BEOG study interviewing staff and coordinate field operations, the continental United States was divided into seven supervisory regions. See Figure 3-2 which shows the supervisory region boundaries as well as the number of sampled institutions in each state. Supervisors managed the interviewers residing in their region and reported to the Field Director as shown in the diagram on the next page.

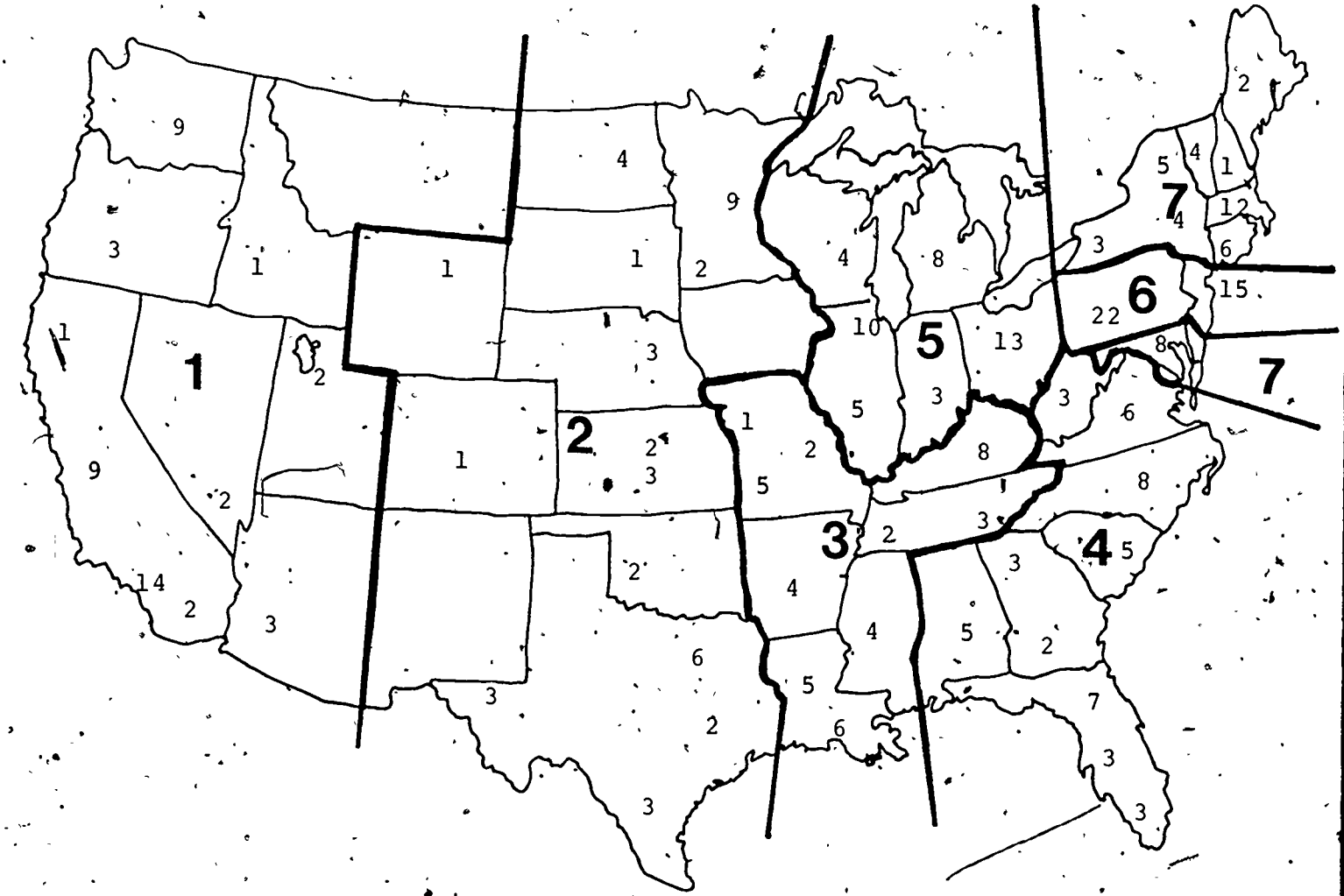
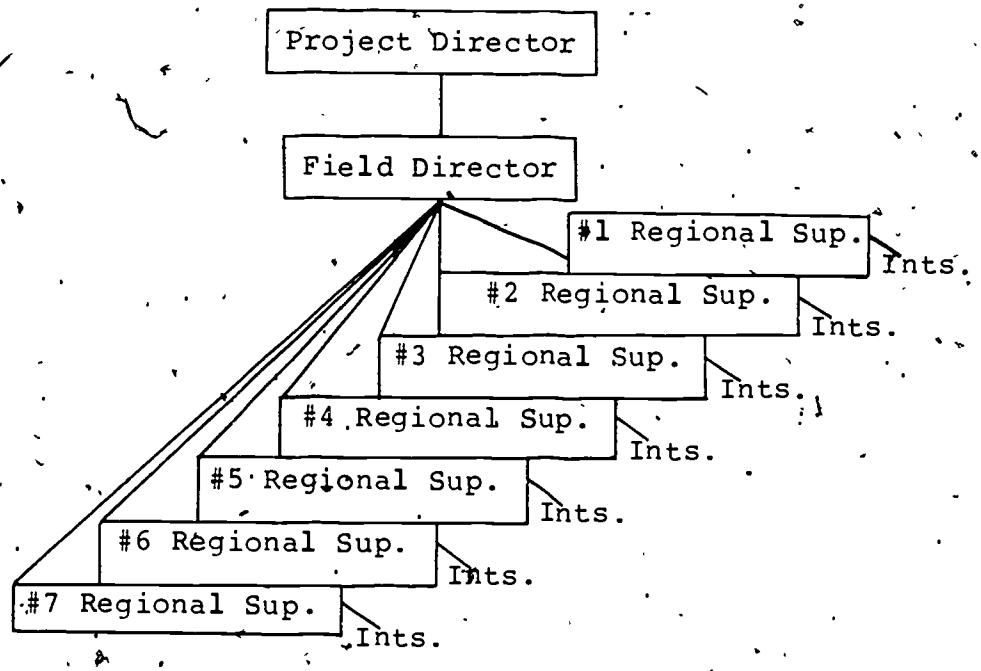


FIGURE 3-2

SUPERVISORY REGIONS FOR STUDENT/PARENT INTERVIEWERS
BY NUMBER OF INSTITUTIONS IN EACH SUBREGION



Interviewer Recruitment

Regional Supervisors and their assistants began interviewer recruitment on December 3, 1980. Primary sources for recruitment included Westat's computerized interviewer file, supervisors' local contacts, local employment agencies, and when necessary, newspaper ads. When interviewer training began on January 26, 1981, 201 interviewers had been recruited and invited to training. An additional 12 interviewers were recruited and trained in February 1981. The major recruitment effort was conducted with minimal information about the location of sample respondents. During December and January the Regional Supervisors knew the location of sample institutions but not the number or location of individual respondents from each institution.

Of the 213 interviewers recruited, 203 attended training and 201 successfully completed training. During training, the interviewers were invited to complete a background information form. A review of these forms shows that 59 percent of the interviewers had worked on at least 1 prior Westat study and 99 percent had previous experience as survey research interviewers.

Training

Training for the seven supervisors and six extra trainers (community leaders) took place the week of January 12, 1981. The first three days were spent acquainting the supervisors and other trainers with BEOG study materials (the questionnaires, field procedures, and interviewer training materials). The last two days were spent training supervisors on their duties (i.e., use of the automated reporting system). Supervisor training was conducted in the Westat offices in Rockville.

The following week was spent training interviewers in the field. The week of January 26, 1981, 109 interviewers were trained in two sites--Cherry Hill, N.J., and San Antonio, Texas. The week of February 2, 1981, a total of 94 interviewers were trained in St. Louis, Missouri, and San Diego, California. Each session was attended by representatives from ED and Advanced Technology, the Westat project director, two lead trainers, supervisors, and other trainers. The trainee/trainer ratio was never more than 1 trainer for each 10 trainees.

Interviewer training for the BEOG survey was based on a training plan which Westat has developed and refined for other

large-scale surveys. Training techniques included home study of a programmed text, interactive lectures to the large group, and role-playing (administration of a mock interview complete with documentation). Central to the training plan is the subdivision of the group of trainees into "learning communities" of approximately 10 trainees who receive individual attention from the "community leader." For the BEOG survey, Field Supervisors functioned in the role of community leaders during training so they were able to observe and assist the interviewers they later supervised. Additional trainers also acted as community leaders and reported the progress of each trainee to their supervisors at the end of each day of training. Trainees who successfully completed a practice interview and the final exam and who received positive evaluations from the comunity leader were given assignments at training.

Field Issues

Traveling Interviewers

Although student respondents were generally located near the sampled institution, parent respondents were scattered throughout the entire country. This meant that at some point, some interviewers would need to travel to the respondents located in far away places. Assignments of this type were held and allowed to accumulate until the last few weeks of the field period. At that point, supervisors coordinated interviewer travel plans with the Field Director, and the interviewers were sent out to conduct the interviews.

Interviewer Attrition

Interviewer attrition on the BEOG study is notable in that it was never a significant problem. Generally, interviewers left the study only after most or all available work was completed in their area of the country.

Interviewer Production

The chart below indicates weekly interviewer production of completed interviews. These totals are taken from Automated Survey Control System reports and represent the number of questionnaires received each week by the Regional Supervisors.

Week #	1	2	3	4	5	6	7	8	9	10
Completes This Week	212	638	1247	1126	1206	1202	1159	812	412	141
Cumulative # Completes	212	850	2097	3223	4429	5631	6790	7602	8014	8155

Reporting Procedures and the Automated Survey Control System [ASCS])

At least twice a week the Field Director contacted the Regional Supervisor by telephone to discuss fieldwork progress and any problems that had arisen. One of these phone conferences was used to discuss the weekly ASCS reports. An integral part of field management on the BEOG study was the computer assisted management system known as ASCS. The ASCS operated through small computer terminals located in the supervisors' homes and connected through telephone lines to a computer. A similar terminal

was located in the home office for use by the Field Director. Each week the supervisors would enter information on field progress into the system, and on a regular basis the terminal would print out summary reports on survey progress. The system was also used to transmit and receive messages to and from the home office, as well as from other Regional Supervisors.

The ASCS generated three reports which were used by the Regional Supervisors. ASCS Report #1, the Supervisor Interviewer Report, listed the I.D. number of all cases currently assigned to an interviewer. Each week a new Report #1 was generated for each interviewer. This report served as a record of assignments and was discussed during the weekly supervisor/interviewer conference. ASCS Report #4 was also generated weekly and presented production information on all interviewers. It provided such things as response rate, hours and expenses per complete, and cost per completed interview. ASCS Report #5 presented totals showing the current disposition of all cases in a region. A review of this report provided an accurate, overall picture of a region's progress in completing the survey.

In addition to these reports, the ASCS was capable of generating 14 more reports for use by home office staff. These reports were used to monitor nationwide survey progress and provided detailed information about different respondent types (e.g., dependent students or dependent parents).

Field Problems

No major problems arose during the field period, and minor issues were discussed with the field staff through the use of Interviewer Field Memos.

SECONDARY DATA COLLECTION

Secondary data collected for the student/parent sample consisted of the following:

- Student Eligibility Reports [SER]
- IRS tax forms
- Financial institution records
- Tax assessor records

Student Eligibility Reports Obtained from Schools

A list of the students selected for study participation was sent to each sampled school, with a request for copies of the SER on file for each of the selected students and a current mailing address and number. Of the 307 schools sampled for participation in the study, 305 schools cooperated in sending copies of the SERs. A total of 4,710 SERs were received from all schools.

When the SER Transcription Log for an institution was complete, it was sent to the data entry office to be keypunched. Using the information keyed from the SER Transcription Log, mailing labels, Call Record labels, questionnaire labels, and mini-labels for filing were generated. In addition, a Master Receipt Control Log was generated to be used for receipt control of the other data sets to be collected during the study.

IRS Tax Forms

Students and parents for whom an address was available were sent a letter of introduction and a package of materials to assist them in preparing for the interview. Included in each of the student packages and in the dependent parent package was a

list of documents they would be asked to show the interviewer. This list was "customized" for each respondent, based on information from the SER. If the SER indicated the parent or the student had filed a 1979 tax return, an "IRS Form 4506, Request for Copy of Tax Form" and an instruction sheet for completing the request were included for the respondent to fill out and return to Westat. Finally, an "information update sheet" was enclosed for the student or parent to fill out with the current names, addresses, and telephone numbers of the student and parent. A pre-addressed postpaid envelope labeled with the respondent's study identification number was enclosed in the package to be used for returning the information update sheet and the Request for Copy of Tax Form. In addition, Request for Copy Forms were obtained from respondents by interviewers at the time of the interview.

IRS Request for Copy Forms returned through the mail were labeled with the study identification number written on the postpaid envelope. The date the forms were received was recorded in the Master Receipt Control Log. These forms were then processed, along with the Request for Copy Forms obtained during the interview, and sent to the appropriate IRS service center.

IRS 1040 and 1040A form photocopies were sent to the receipt control office by the IRS Service Centers in packages containing an invoice, a list of the photocopies contained in the shipment, a list of the requests unfilled because the service center was not able to locate the form, and a list of requests unfilled because the service center had no record of the requested form.

Financial Institution Records

A Financial Institution Authorization to Release Information Form was completed in the field for each account maintained by a respondent who claimed to have more than $4,000 in checking and savings accounts at the time of the BEOG application. The completed authorization forms were removed from the questionnaire booklet at the time the case was processed through receipt control.

Two copies were made of each release form. The original release and one copy were sent to the manager of the financial institution named by the respondent, along with a covering letter explaining the study and the participation required from the bank manager. Each financial institution was asked to ascertain the respondent's account balance as of the date of application for a Basic grant and to record the balance on the release form. The release form was then to be returned to the receipt control office in a postpaid envelope.

A total of 422 completed Authorizations to Release Information were obtained from 302 respondents reporting more than $3,999 in checking and savings accounts and sent to financial institution managers. The managers were requested to report the respondent's account balance on the date of application. The 375 completed returned forms provided financial information on 270 respondents.

Tax Assessor Record Study Results

A total of 78 schools, or 25 percent of the schools in the sample, were systematically selected after being stratified by

size and type for inclusion in a study of tax assessors records of home value. All homeowners in the 78 schools were selected for study. Of the 1,260 sampled applications in the 78 schools, 568, or 45.1 percent, were dependent or independent student homeowners.

Questionnaires requesting reports of the most recently assessed market value were sent to local tax assessors based on the current address recorded on the Student Eligibility Report. The assessors were to return the completed questionnaires to the receipt control office in the postpaid envelope provided. Telephone follow-up was used to prompt late responders. A total of 466 assessors responded with usable data.

INSTITUTIONAL DATA COLLECTION

This section describes all aspects of the institutional data collection, including:

- Development and field testing of the data collection instruments
- Site visit scheduling
- Interviewer recruitment and training
- Field procedure
- Quality control procedures and field supervision
- Post data collection debriefing

The data collection instruments used in the institutional component of the study were developed during the first weeks of the study from the instruments used in the previous BEOG Quality Control study. Building upon the experience of the earlier

study, the Institutional Interview Questionnaire [IQ] and the Student Record Abstract [SRA] were revised and refined. Once the final versions of the instruments were produced, they were submitted for Federal forms clearance.

A series of field tests played an integral part in the development of the instruments. The results of the field tests were used to (1) revise and improve the instruments, (2) establish field data collection procedures, and (3) develop interviewer training materials.

Nine institutions located in the Washington metropolitan area and roughly representative of the larger sample were contacted to participate in the field tests. Figure 3-3 shows these institutions and their characteristics.

The field tests were conducted by senior project staff in two-person teams. One member conducted the interview and the second recorded comments regarding responses to the items, difficulty in understanding the questions, the order of the questions, questions demanding further probing by the interviewer, or any other observations that might have helped in the revision of the instrument. Following the interview, the FAA was asked to critique both the questions and the interviewing techniques used.

The financial aid files of nine Basic Grant recipients selected at random by the FAA were reviewed, and their information was recorded in the SRAs. The Bursar and Registrar were visited if financial or registration information was not

INSTITUTION	CHARACTERISTICS
Prince Georges Community College	2-year/public
Northern Virginia Community College	2-year/public
Montgomery College	2-year/public
Catholic University	4-year/private
George Washington University	4-year/private
Howard University	4-year/private
George Mason University	4-year/public
Strayer Business College	Less than 2-year/proprietary
Control Data Institute	Less than 2-year/proprietary

FIGURE 3-3

INSTITUTIONS PARTICIPATING IN PILOT
TESTING OF INSTITUTIONAL DATA COLLECTION INSTRUMENTS

available in the financial aid office. While at the institutions, detailed notes were taken on the effectiveness of the Student Record Abstract, the length of time of each file review, and any problems that might be encountered in locating student financial aid data.

Site Visit Scheduling

In designing the field organization and travel itineraries, every effort was made to balance the need to minimize travel expenses with the need to allow sufficient time for data collection, travel, and rest. After contacting and gaining compliance from all 305 institutions, a firm site visit schedule--including airline, hotel, and car rental arrangements--was established. Interviewers were required to adhere to this preestablished itinerary as closely as possible. Each interviewer's field schedule was monitored closely by the project office throughout the eight weeks of data collection. A step-by-step description of the procedures used to establish a site visit schedule follows.

1. After field testing the survey instruments and the field procedure, we estimated that an interviewer could complete 12 file reviews a day. Based on this knowledge of what constituted an average work day, the number of work days for each of the 305 institutions in the sample was calculated.

2. The map of the continental United States was divided into 13 contiguous regions, each with approximately the same number of work days, travel days, and rest days. Every effort was made to capture "clusters" of institutions within each region. The

average region contained 24 institutions (more where institutions were relatively dense, such as the mid-Atlantic states, fewer where institions were relatively sparse, such as the Rocky Mountain states).

3. A tentative site visit schedule for each region was established with travel routes within each region designed to minimize travel expenses.

4. Advanced Technology contacted each of the 305 institutions at least 3 times prior to the site visits. In January 1981, two months before the data collection began, letters were sent to all institutions describing the study's objectives and asking for participation. Telephone calls followed two weeks later. During these calls, senior project staff followed a defined protocol. They:

- Asked if the FAA had received and read the initial letter
- Reviewed the purpose and requirements of the data collection effort
- Informed the FAA of the tentative site visit dates
- Asked if the dates were acceptable
- Asked for acceptable alternate dates if the initial site visit dates were unacceptable
- Asked for any specific information that would simplify the interviewer's task (i.e., directions to the aid office, parking arrangements, and location of the Bursar and Registrar)

5. Following the telephone calls, a firm site visit schedule was established and letters were sent to all institutions confirming appointment times.

Interviewer Recruitment and Training

Advanced Technology advertised for interviewers in major city newspapers, the Chronicle of Higher Education, and the NASFAA Newsletter. Nearly 100 resumes were received and reviewed; 36 individuals were interviewed; and 13 were hired. All 13 had considerable financial aid experience; all had bachelor's degrees; and 3 had advanced degrees. We felt that interviewers with student aid experience would be far more effective data collectors than those without such experience. Experienced student aid administrators would be able to effectively probe for answers from FAAs; they would be able to recognize significant answers; they would know when impressive sounding jargon was actually vague or meaningless; and they would collect data on individual students quickly and efficiently given their familiarity with the record keeping practices of institutions of higher education.

Two manuals were developed for training. The Interviewer Training Manual describes the study and outlines precisely the procedures to be followed at each site, including information on:

- Recording student record data
- Editing
- Returning forms to the project office
- Travel logistics

The Question-by-Question Specifications Manual describes each question in the interview and record review form, giving information on the various responses to expect.

The interviewers were trained the week of March 23, 1981, in McLean, Virginia, in all the data collection activities they subsequently pursued in their field work. In addition, they received training related to reporting, scheduling, expense reports, and mailing and receiving survey forms. The training included field practice at seven Washington, D.C., metropolitan area institutions.

A day-by-day description of the principal training activities follows:

Monday morning	Review of the background of the BEOG Quality Control study and Campus-based programs, and summary of the interviewers' tasks and responsibilities
afternoon	Introduction to Institutional Interview Form; presentation of film: Introduction to Interviewing; demonstration of institutional interview
Tuesday morning	Role playing of institutional interview with trainees in pairs; item-by-item examination of Student Record Abstract [SRA]
afternoon	Completion of four SRAs using hypothetical student files of varying complexity; explanation of Alternate Disbursement System
Wednesday morning	Role playing of resolution/exit interview with trainees in pairs; explanation and practice with the Corrections Control Group forms
afternoon	Explanation of shipping and receiving of data collection materials, field editing methodology, and travel arrangements; explanation of data receipt and processing system

Thursday	Field practice at seven metropolitan Washington institutions
Friday	Detailed review of field practice; completion of paperwork

Field Procedures

This section provides a summary of the procedures used to conduct the institutional interview and to administer the survey instruments. (For a more detailed discussion of the interviewing protocol and the procedures used to complete the SRAs and CCGs, refer to Advanced Technology's Interview Training Manual). Figure 3-4 illustrates the complete data collection cycle discussed in this section, from the confirmation of the interview time to the sending of completed forms to the project office. A step-by-step explanation of the procedures in this cycle follows:

1. At least two days--but not more than one week--prior to the site visit, the interviewers called the FAA to confirm their appointment time and to get instructions for parking and locating the financial aid office.

2. After arriving at the institution, and before administering the interview, the interviewers met with the FAA to introduce themselves and the study. The following are items that were covered in the introduction:

- A description of the purpose and nature of the study, with particular emphasis on the goals of the institutional component of the study. Even though they had received letters and phone calls, many FAAs had only a sketchy idea of the study prior to the site visits.

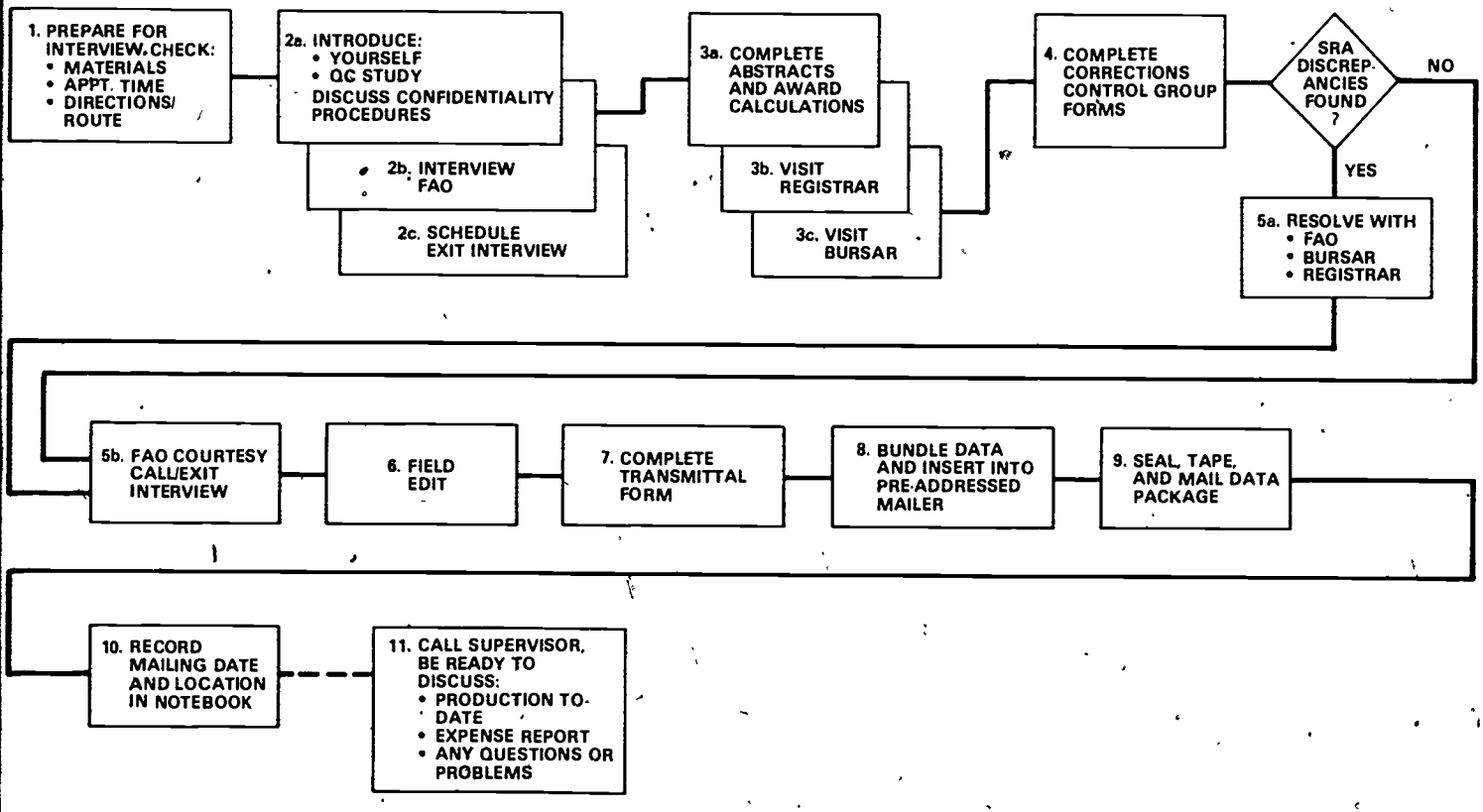


FIGURE 3-4

INSTITUTIONAL DATA COLLECTION CYCLE

- An assurance of the confidential nature of the study. The interviewers read the "Confidentiality Statement" found on the cover of the IQ and, if it seemed that the FAA was particularly sensitive about the confidentiality issue, explained the various measures employed during the data collection and data processing stages of the study to maintain confidentiality.

- A brief step-by-step itinerary of the data collection activities at the institution, describing the general purpose of each step.

3. The interview, normally lasting from 60 to 90 minutes, followed. To assure complete, accurate, and consistent data, the interviewers were instructed to observe the following basic guidelines when asking questions and recording responses:

- Remain neutral
- Ask all questions exactly as worded
- Discourage unrelated conversation
- Ask respondents to enlarge or clarify answers when necessary
- Record verbatim the FAAs' responses

At the conclusion of the interview, a follow-up or exit interview was scheduled with the FAA.

4. After the initial interview with the FAA, the interviewers reviewed student files and recorded the information in the SRA. A notice stating the purpose and date of the data collection was placed in each reviewed file.

5. Before leaving a site, the interviewers conducted a brief exit interview with the FAA. The purpose of the exit interview was to thank the FAA for his or her cooperation and to

discuss SRA discrepancies to learn whether the FAA could offer a logical explanation for what on the surface appeared to be an error or violation of the BEOG program. The interviewers were instructed to use tact in seeking these explanations, reassuring the FAAs, if necessary, that their responses would be kept in confidence and used for national estimates only.

6. At the end of each workday, all completed instruments were reviewed for possible omissions, inconsistencies, illegible handwriting, or misplaced codes. If interviewers were scheduled to return to the institution the following day they could take advantage of this opportunity to clarify or retrieve any missing information. Otherwise, interviewers did not recontact institutions for data retrieval purposes.

7. The interviewers were instructed to mail survey data to the project office every two or three workdays or before a weekend or extended travel period. The interviewers recorded on a transmittal form the name of the institution(s), the number of specific items being returned in the mailer, and all other information that explained the status of the returned data. The transmittal form and the survey forms were secured with a rubber band and placed in a pre-addressed, business reply mailer. Upon receipt at the project office, the contents of the mailed packages were verified against in-house records of interviewer assignments. (See Chapter 4 for a further discussion of receipt procedures.)

Data Collection Supervision and Quality Control

In order to properly conduct a field data collection of this magnitude, it was necessary to establish well-defined procedures to ensure:

- The quality of survey data
- The confidentiality of survey data
- Full communication between field and project office staff
- The orderly flow of survey materials between the project office and the field interviewers

Quality Control

Rigorous quality control procedures were required to ensure the validity and reliability of the collected data. These included project office review of the data, project office telephone validation, and on-site observation of interviewers by project staff.

1. Project Office Edits

After the instruments arrived in the project office they were scrutinized by a series of manual and computerized edits. When critical omissions or ambiguities were discovered in the questionnaires, the interviewer was contacted immediately for an explanation. Occasionally an institution was contacted directly if survey data needed clarifying. A detailed description of the project office's coding and editing system can be found in Chapter 4.

2. Project Office Validation

Project office validation verified that the institutional data collection was, in fact, conducted according to correct procedures. Each week all institutions reported as having been visited the previous week were telephoned to confirm that the FAA was interviewed, that the student records were inspected, and that the conduct of the interviewer was appropriate. In addition to this general performance validation, one institution was selected from among those that each interviewer had visited that week. That institution was asked to verify two items on the IQ and one item on a randomly selected SRA. The items selected for validation were those unlikely to have changed in fact or in the respondent's perception between the time of the site visit and the validation call. The findings of each call were carefully documented. Institutions where data collection was observed by a home office staff member were not called for data validation. If an FAA had made a negative evaluation or a discrepancy had been discovered between the interviewer's findings and the validator's findings, the interviewer would have been contacted immediately for an explanation. Virtually all the FAAs called, however, gave very positive evaluations of the interviewers' performance during the site visits. In addition, no data discrepancies were found.

3. Field Observations

Field visits allowed the project staff to observe first hand the data collectors' interviewing techniques, professional manner

and thoroughness. A report was written for each site visit, identifying areas of weakness and noncompliance with approved procedures. Soon after the interview and file reviews--and while still on-site--the field monitor reviewed each item on the report with the interviewer, noting strengths and weaknesses. If areas of improvement were identified, the field monitor added specific suggestions for improvement on the field report. Each interviewer was observed twice: once during the second week of the data collection and once during the sixth week. After each site visit, the field monitors met to discuss their findings. If general areas of improvement were identified, they were noted in the next memorandum sent to all field interviewers.

Confidentiality Procedures

To protect the confidentiality of the survey data, the following procedures were employed during the data collection:

- This statement (found on the front of every IQ) was read to FAAs before each interview.

> CONFIDENTIALITY AND PRIVACY
>
> This study is being conducted according to the regulations and provisions of Subsection (e) (3) of the Privacy Act of 1974, as amended. The information (I/we) collect at your institution will become part of the existing BEOG system of records; however, that data will be aggregated in such a way as to make identification of a particular institution's records impossible. (I/we) have signed a confidentiality statement and, except for the express purpose of this study, (I/we) have sworn not to reveal any information you give (me/us) during this interview or from (my/our) review of your student files, except as required by law.

- All interviewers signed an assurance of confidentiality statement.
- All interviewers kept completely confidential the names of respondents, all information or opinions collected in the course of interviews, and any information about respondents learned incidentally.
- Survey data containing personal identifiers were kept in a locked container or a locked room when not being used each working day in routine survey activities.

Communication and Reporting

Full communication between field and project staff was crucial to ensuring quality data and maintaining the tight field schedule called for in the survey. Supervision of the data collection effort took place primarily through scheduled weekly telephone calls from the field staff to the project office. Periodic memoranda served as a means for informing field staff of any update to or changes in data collection procedures.

1. Telephone Procedure

A separate telephone line with an 800 number and a recording device was installed in the project office. Each week on a scheduled day and time interviewers were required to call the project office. These weekly calls served three purposes:

- Monitoring of Data Collection

  The calls provided an opportunity for the data collection manager to review with the interviewer any problems or error patterns identified by the coding and editing staff. (See Chapter 4 for further discussion of this procedure.) The data collection manager during this time answered questions regarding interviewing procedure, survey form administration, and coding convention. These weekly calls also gave the data collection manager the opportunity to discuss the contents of the field memoranda with the interviewers.

- Changes in Interview, Travel, and Accommodation Arrangements

  Every effort was made prior to the field data collection to establish a schedule that would allow sufficient time for interviewing, file reviews, travel, rest, and meals. Often, however, there were unforeseen changes in the interviewers' itinerary initiated by FAAs, hotels, airlines, or the interviewers themselves. In all cases the interviewers were required to report schedule changes to the project office. Since the itineraries were arranged centrally at the project office, the data collection manager was frequently informed about travel arrangement changes prior to the field staff. In these cases, the data collection manager informed the interviewers about schedule changes. If the progress of a particular interviewer was found to lag due to unforeseen data collection or travel problems, the data collection manager discussed with the interviewers ways of resolving the schedule problems during the weekly calls. Sites of interviewers whose progress was lagging were often reassigned to other interviewers who were ahead of schedule. In emergencies, project office staff were sent to the field to aid interviewers with schedule problems.

- Clarification of Routine Business Matters

  Issues related to expense reports, travel advances, paychecks, and mailing procedures were also discussed during the weekly calls.

The interviewers were encouraged to call the project office more frequently than the required weekly call. A recording device was installed to take messages after business hours, and all interviewers were given the telephone number of a staff member to call in case of an emergency.

2. Field Memoranda

Field memoranda were issued to communicate updates to existing procedures or implementation of new procedures. To ensure that the interviewers understood each item of the memos, the data

collection manager discussed the memos with the interviewers during the weekly calls.

Survey Materials Management

A mailing system was established to ensure that the interviewers were adequately supplied with survey materials during the eight weeks of data collection. Materials were mailed to field personnel on a periodic basis throughout the data collection effort. Well in advance of site visits a separate mailing schedule was established for each of the interviewers. These mailing schedules were used to determine the number of institutions each package would cover, the exact number of all instruments required for these institutions, and the institution most appropriate to receive and hold the package for the interviewer.

The packages were mailed three weeks prior to the scheduled site visit. One week prior to the scheduled site visit the institution was called to verify the arrival of the package. If the package had not arrived, a duplicate package was prepared and sent by express mail.

Interviewer Debriefing

The debriefing of the institutional interviewers, held on May 23, 1981, was an integral part of the overall data collection effort. In attendance were the 13 interviewers, all the home project staff, and the Project Officer. Each of the interviewers had considerable student financial aid experience and, therefore, had much to offer during the day-long discussion in the way of

observations and recommendations regarding the Basic Grant delivery system.

The debriefing was organized into five sessions designed to focus the discussion and assure that all pertinent issues were addressed. A summary of the principal debriefing topics follows.

- BEOG Delivery at the Institutional Level

  The interviewers compared institutional procedures observed during the data collection, discussed institutional validation, recalled significant problems with Basic Grant delivery noted at institutions, and made recommendations for improvement to the Basic Grant program.

- Evaluation of Data Collection Instruments

  The interviewers singled out questions that needed improvement.

- Evaluation of Training Program

  The interviewers discussed how well the training had prepared them for their field experiences and noted areas where instruction or greater emphasis was needed.

- Status Report on Analysis of Institutional Data

  A senior staff member gave a short overview of the major findings to date.

- Evaluation of Logistical Support

  The interviewers evaluated the itineraries, telephoning procedure, field memoranda, field observation, hotel and travel arrangements, cash handling procedures, and general staff support.

CHAPTER 4

DATA PROCESSING

The first section of this chapter describes the procedures used by Westat to compile, edit, and enter student, parent and secondary data. The second section discusses the procedures used by Advanced Technology to process institutional data. The final section provides a technical discussion of the final data file merge.

STUDENT/PARENT

Six primary data sets were prepared for the student/parent component of the BEOG Quality Control study. The preparation of each data set required the application of similar data preparation procedures. These procedures are described in this section of the report.

Receipt Control

The function of receipt control was to provide a catalogue of all data documents received by the Coding Office of the Basic Grant study. The receipt control procedure was slightly different for each of the data sets.

As SERs were received from the sample schools they were logged in on a computer printed listing of the students sampled from the schools. Data from the SERs and data provided by the schools were used to produce a label file of names, addresses, and Social Security numbers of all sampled students, and the

names and addresses of their parents. This label file was used to produce a master receipt control log for cataloguing Student and Parent Questionnaires and IRS forms returned from the field.

The master receipt control log listed the study identification number, name, address, and Social Security number of each student/parent pair. The log was organized in numerical order by study identifier within school. Space was provided for recording updated information on names and addresses, the date each questionnaire was received from the field, the completion status of the document, and the coding batch number assigned to the document. Labeled columns were also provided for recording the receipt of IRS Release Forms by mail (from the initial information mailing) and inside the questionnaires (obtained during the interview). When IRS 1040 and 1040A Forms were received from the IRS Service Centers, they were coded with case identifiers and logged in the master log, with the IRS coding batch number.

Coding

A coding manual was prepared for the Basic Grant study for use in training the data preparation staff; serving as a complete and detailed reference for analysts, programmers, and data preparation staff; and providing documentation for the Basic Grant study data files. The coding manual consists of an introduction to the study procedures and purposes, a review of general data preparation procedures to be followed, and coding and editing specifications for each of the six data sets.

Approximately 40 survey processing personnel were selected for training as coding and editing staff for the Basic Grant study. Four group leaders were selected from this number based on their skills and qualifications. The group leaders were assigned as assistants to the coding supervisors and as coder verifiers. As much as possible, experienced Westat survey processing personnel were selected as coders for this project in order to minimize the amount of training necessary on basic coding skills.

Coders were trained in groups of between 4 and 14. Each group was trained to code on one of the six study data sets. Between one and eight hours of training time were required, depending on which data set was to be coded.

Questionnaires were precolumned before printing so that coding could be written directly on the questionnaires. SERs and IRS 1040 and 1040A Forms could not be precolumned, so transcription sheets were designed for the coding.

Coders were assigned work by coding batch and were required to complete the coding of one batch before beginning work on another. Errors found during verification by the supervisor were discussed with the coder committing them. If persistent errors were discovered, a coder would be asked to review previous batches and correct them. Problems found during coding but not resolved in the coding specifications were documented and referred to a supervisor to be resolved.

Occasional problems with illegible figures arose in the coding of photocopies of Student Eligibility Reports and IRS tax forms. It was necessary to code illegible data elements as missing values in these situations.

The major coding problem for the Student and Parent Questionnaires was the large number of questions in each questionnaire which were open-ended in format. It was not possible prior to the beginning of coding to devise lists of all the possible responses to these items. Since this problem was expected, a controlled system of dealing with it was implemented at the beginning of coding: Responses which were not codable in the predetermined list of codes in the coding specifications were documented and referred to the supervisors who constructed codes for the new items. New codes were published each morning on a Coding Change Sheet. Coders were responsible for keeping their manuals up to date and were required to record each issue of the coding changes in a log.

Data Retrieval

Coders were trained to "edit" the data collection instrument during the coding. The edit function involved checking for readability, sensibility, and following of skip patterns. (The editing function was much more important in the coding of the questionnaires than in the coding of the secondary data sources.) A general rule was established that all primary verification questions in the questionnaire must have codable responses. When coders found erroneous skips, illegible answers, or illogical

responses in any of the verification questions, they documented the problem and referred the case to a supervisor for data retrieval.

Experienced, specially trained telephone interviewers were used for telephone data retrieval of the problem items. Case problems were described on a Data Retrieval Request Form which also served as a Record of Calls for the interviewer. Data retrieval was attempted on 201 Parent Questionnaires and 139 Student Questionnaires.

In addition to data retrieval requests due to problems found in the coding edit, data retrieval requests were also generated during machine editing. In fact, a substantial portion of the data retrieval requested on the Parent Questionnaires occurred during an attempt to obtain more reasonable estimates of the value of assets and debts on assets from respondents.

Key Data Entry and Machine Editing

Coded documents and questionnaires which had been verified were transmitted to the Westat data entry staff in groups (called "keying batches") of approximately 100 documents. Coded documents were keyed into an in-house disk storage system, and then key verified from the disk. After keying and key verifying, the data were transmitted via telephone link to the main computer where they were stored on tapes to await machine editing.

All data sets except the Tax Assessor File and the Financial Institutions File were machine edited with special purpose COBOL programs, written to check for out-of-range codes, incorrect skip

patterns, and inconsistent response patterns. The data sets were edited by keying batch.

Machine edit staff were trained coders, the majority of whom had previous Westat experience as machine edit clerks. A supervisor with machine editing experience was assigned to oversee the machine editing of the questionnaire files.

File updating instructions were written on transcription sheets by the machine edit clerks, checked by the machine edit supervisor, and then sent to the data entry office for keying and transmittal to the computer center. Updates were made to the files by a special purpose COBOL update program. After each update run was complete, another editing cycle was run to verify that corrections had been made and to check for new errors. The update-edit cycle was repeated until each batch of data in the data set was clean.

When data sets were complete and cleaned in the machine edit process, a cross-file merge to check for missing cases and inconsistent student status (dependent/independent) matches was performed. Mismatches between internal status (based on answers to questions in the questionnaire) and external status (based on information from the school provided copy of the Student Eligibility Report) were carefully checked for accuracy. Mismatches between the internal status of the Student Questionnaire and the internal status of the Parent Questionnaire were also checked.

Where true mismatches of status occurred, the original status of the questionnaires was altered to indicate the final dependency status of the case.

In addition to the merge to compare status codes, a comparison of the names of dependents listed on the parents' income tax forms to the names of the students was made. A data set consisting of the study identification number and an indicator of the search result was produced.

Frequency distributions were run on all variables in each data set after the machine edit process was complete. The frequency distribution was proofed for any inconsistencies and errors not found in the machine edit process.

INSTITUTION

This section describes the procedures used to process and verify institutional survey data from their receipt at Advanced Technology to final entry into a computer data base. The survey form processing system described in this section included quality control procedures designed to:

- Maintain the confidentiality of all survey data
- Ensure the accuracy of data provided by the institutional field interviewers
- Ensure the transcription accuracy of coding clerks and keypunchers
- Maintain control of the status of all data collection instruments in order to minimize the possibility that instruments would be lost

Confidentiality of Institutional Survey Data

Several procedures were instituted to assure the confidentiality of survey data. All incoming packages were maintained in locked cabinets in a locked storage area. Access to the survey data was limited to those persons working on the BEOG Quality

Control study and to those who had been instructed in the study's confidentiality requirements. As with the field interviewers, all Advanced Technology data entry personnel were required to sign confidentiality pledges.

Receipt Control

The successful completion of the project, given the large number of data collection instruments, required a well-defined system to be used by clerks in the receipt, logging, and routing of all institutional data received from the field interviewers. The following procedures were used to maintain control of the status of all cases still out in the field as well as the location of every form received in the home office:

1. All incoming packages of instruments contained a transmittal form describing in detail the contents of each package. If a discrepancy was noted between the actual contents of the package and its corresponding transmittal form, the interviewer was contacted immediately for an explanation.

2. The contents of each package were checked against a master list of the interviewer's institutional assignments. If it appeared that an assignment had not been completed, the interviewer was contacted.

3. Once it was determined that a complete institutional assignment had been received at the project office, the instruments were sorted into reference groups of manageable size. Each reference group, or package, was assigned an ID number which was recorded in a master control log. The master control log was

used to trace the path of each instrument through all the processing steps. It served as an excellent mechanism for maintaining control over the status and location of the survey data. Each line entry of the log identified an instrument package by number and described the exact status of that package.

Manual Editing and Coding

A coding staff of five under the close supervision of a senior coder thoroughly reviewed each survey form for completeness and accuracy. The coders checked skip patterns to see that they were followed correctly, checked responses for clarity and relevance, and checked for the consistency and logic of all data. Although the coders were instructed to scan all items on the survey forms, certain questions were found to be error prone and, therefore, were given particular attention.

A coding and editing specifications manual, developed by senior project staff prior to the receipt of data, guided the coders. The manual was used to train the coding staff and served as a detailed reference for analysts and programmers. The manual included:

- A summary of the study
- An explanation of general coding and editing procedures
- Question-by-question instructions
- A list of error prone questions

Often missing or incorrect data could be reconstructed on the basis of responses to associated questions, interviewer

notes in the margins of the survey form, or information from current school catalogues and financial aid materials.

If significant omissions or errors were discovered interviewers were contacted by the data collection manager. In the rare instances when interviewers could not provide the needed data, the institutions were contacted directly. During the eight-week data collection there was never a need to revisit an institution to retrieve data.

A procedure was established to record and inform interviewers of noncritical recording errors. A log was used to record problems and suggestions for improvement for each interviewer. At the end of each week the logs were routed to the institutional data collection manager who forwarded the information to the interviewers during their weekly telephone calls.

Following the preliminary edit, the survey form data were coded for entry into a computerized data base. Often a coding situation arose that was not directly addressed by the coding and editing specifications manual. If a situation could not be adequately resolved by existing coding convention or if a response could not be coded with any of the provided codes, the coders were directed to record the situation on a form and refer it to the senior coder. The senior coder, in consultation with other BEOG Quality Control analysts, if necessary, made a decision on each referred case and recorded the decision directly onto the "problem" form. The problem forms were then filed by question number and served as a permanent record of all coding decisions. Often during the coding process it was necessary to update the

coding and editing specifications manual with new codes, changes in data ranges, and changes in coding convention. It was extremely important that all coders be informed of such additions and changes. As a rule, at the end of each workday, a memo was distributed to all coders detailing changes. The coders would then incorporate the changes into their manuals.

Once completely coded and edited, instrument packages were routed to the coding supervisor for verification. At the beginning of the coding process all survey forms were scrutinized for coding errors. If a pattern of errors was noted that coder was retrained in the deficient area. Once an individual coder reached an acceptable level of accuracy, a sample of that coder's package was inspected. In addition to the above item-by-item verification, certain items which proved to be particularly error prone were verified 100 percent in all instruments. For example, the questions in the Student Record Abstract pertaining to institutional error were verified thoroughly.

Keypunching

In the early stages of the project, six data entry firms were contacted. We decided to engage Data 1 Associates of Springfield, Virginia, because of its proven ability to produce high quality work with speed and economy.

Detailed keypunching instructions were forwarded to Data 1. The first keypunched packages returned to Advanced Technology were examined closely to verify that keypunch instructions were being adhered to. Keying was performed in packages, transferring data directly from instruments to cards. Each package of keyed

data was 100 percent key verified against the associated instruments. Key verifying was performed by a key operator other than the one who performed the keying.

To maintain control over the status of all survey forms, a well-defined procedure was established to transmit the data to and from Data 1. A weekly standard time for pickup and delivery was set and maintained. A transmittal form detailing the contents of each package of instruments accompanied each delivery. The signature of a Data 1 official on this form verified the totals of each delivery. Once returned from keypunching, each package was checked for completeness and filed.

Machine Editing and Updating

As keypunched cards were returned from Data 1 they were entered on a data file and subjected to a series of computerized edits. The purpose of the machine edits was to (1) act as a backup of the manual editing, (2) perform edits that would be very time consuming if done manually, and (3) discover keypunching errors.

Under the guidance of a senior analyst a series of EASY-TRIEVE programs were designed to:

- Check for missing data
- Check data ranges.
- Check for incorrect skip patterns
- Check consistency between data items

The edit programs were tested on the first set of keypunched data. The results of these tests were reviewed by senior staff to ensure that the edits had the proper stringency.

Edit failures were individually listed for resolution by the coding staff. The original data collection instrument was reviewed, and file updating instructions were written on the error listing. The data file was updated and a new edit cycle was run. This update-edit cycle was repeated until no data imperfections remained.

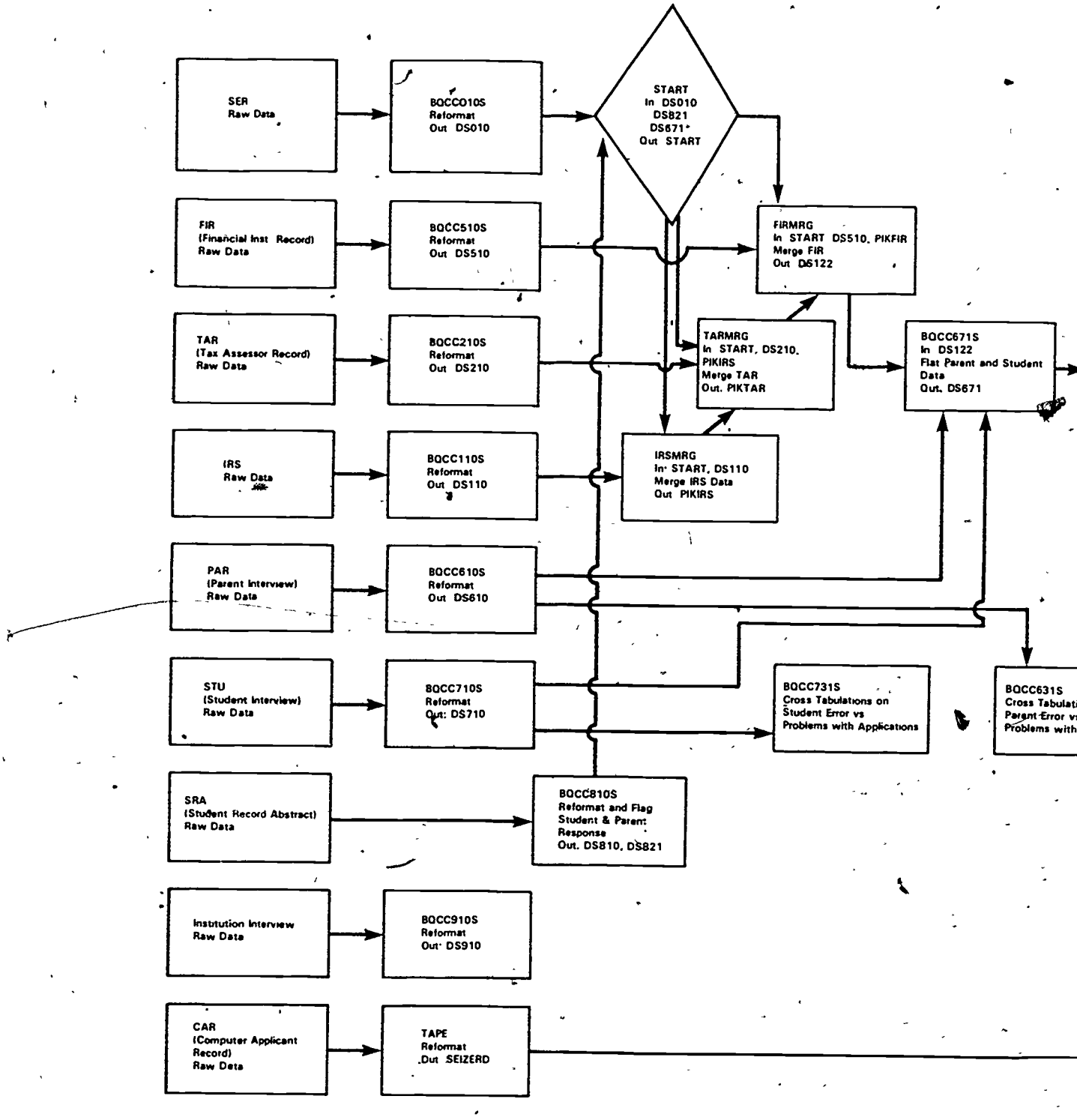
Frequency distributions were run on each variable once the editing process was complete. The distributions were inspected by senior staff for errors not detected by the manual or machine edits. If errors were found, the original instrument was reviewed and the file updated.

SCHEDULE OF DATA MERGE

The first step in the data merge was reformatting the clean raw data tapes into SAS files. The following programs were run:

- BQCC010S was run on the SER data.
- BQCC110S was run on the IRS data.
- BQCC210S was run on the TAR (Tax Assessor Record) data.
- BQCC510S was run on the FIR (Financial Institution Record) data.
- BQCC610S was run on the parent data.
- BQCC710S was run on the student data.
- BQCC810S was run on the SRA (Student Record Abstract) data.
- BQCC901S was run on the institutional interview data.

All of the SAS programs were stored on Librarian modules. They were accessed and executed by short JCL sequences stored in an active COMNET workspace.



*DS671 Used In Program START Was Created In A Preliminary Merge Which Omitted Programs START, IRSMRG TARMRG FIRMRG

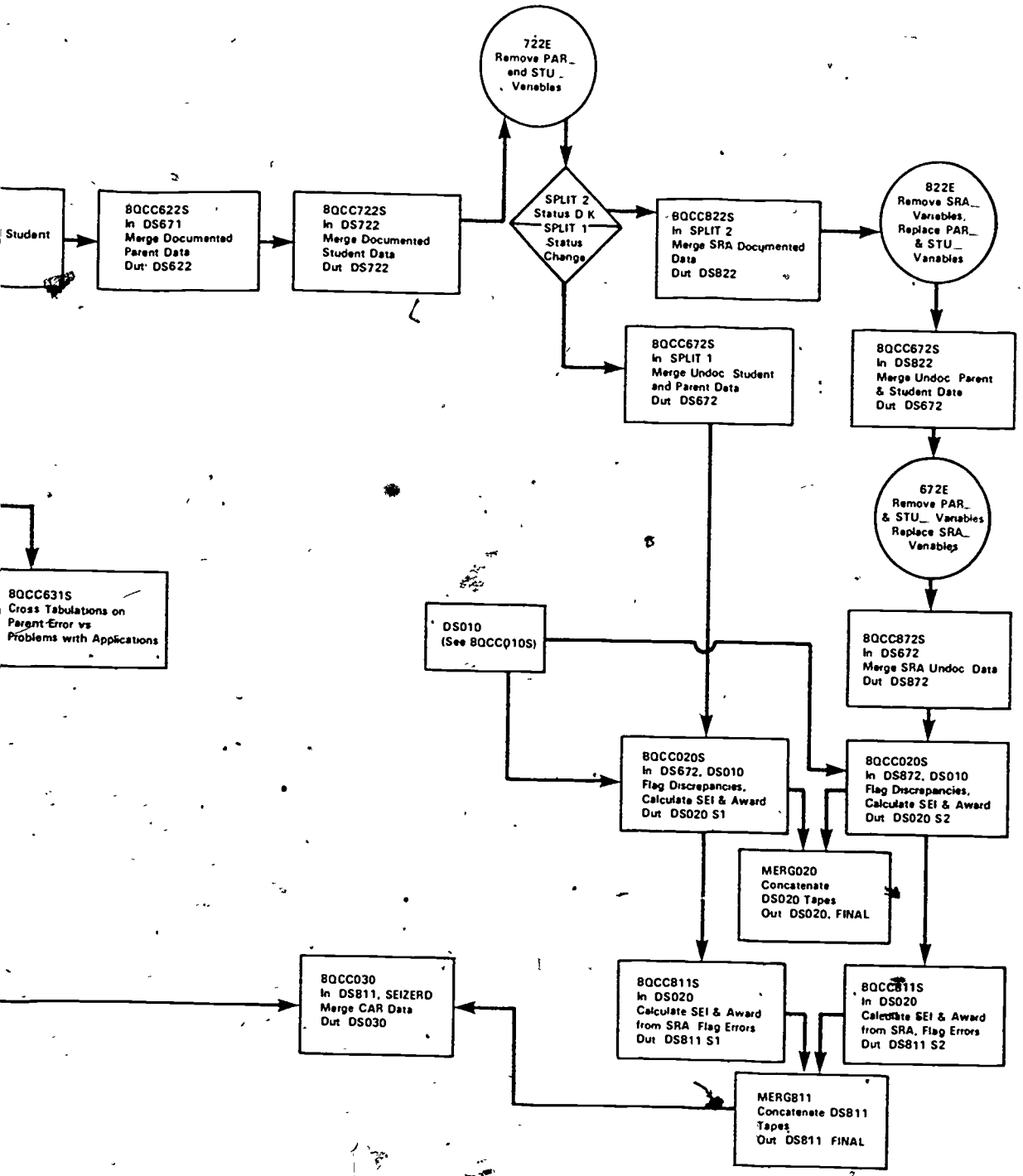


FIGURE 4-1

FLOWCHART OF DATA MERGE

The naming convention for tapes created by Data Merge programs was BGP.BQC.DS___, where the blank was filled by the three-digit number corresponding to the program in which it was created. Thus, program BQCC010S created tape BGP.BQC,DS010, etc. The next step was to run a program called START. This program reads a computer applicant record tape and a tape containing student, SRA, and parent data flags. These flags, created in earlier runs of the data merge, indicate what information existed for each ID. The program START assigned each ID a flag to code which IRS information was needed, according to dependency and marital status. This information was stored on a tape named BGP.BQC.START319.

Using BGP.BQC.START319 and the IRS data tape, the merge was begun by a program named IRSMRG. This program picked out the appropriate IRS data for each ID according to the code from tape BGP.BQC.START319 and created flag variables to indicate whether IRS data had been found for each item. The resulting tape was PIKIRSK1. Similar programs TARMRG and FIRMRG were run to collect tax assessor and financial institution data. Tapes created were BGP.BQC.PIKIRSK1, BGP.BQC.PIKTARK1, and BGP.BQC.DS122, in that order. The program which cataloged BGP.BQC.DS122 created the "best value" variables, to be filled through the merge. The merge continued with program BQCC671S, which created flags to determine the source and documentation of the values from student and parent files. This program read tape BGP.BQCDS122 and the universe file UNIVINT3 and created tape BGP.BQC.D671. The next

program in the sequence was BQCC622 which merged the documented parent data. It read tape BGP.BQC.DS671 and created BGP.BQC.DS622. Then the student data were merged by program BQCC722S, which read tape BGP.BQC.DS622 and created BGP.BQC.DS722. At this point a short program was run to delete the SQ___ and PAR___ variables from the student and parent files. The tape created still retained the best value and flag variables, but the other variables were dropped to save space for the next step in the merge.

At this point two programs were run to split the observations into two groups: those students whose dependency or marital status was determined by us to have been incorrectly reported on the SER, and those students for whom the status items were correct there. The programs were named SPLIT1 and SPLIT2, respectively, and created tapes named BGP.BQC.DS722 (differentiated by a fourth-level name). This split was necessary because for the first group all SER and SRA data were incorrect, while for the second group the SER data might still have been the best available data. Thus, these groups had to be split before merging the SRA data.

A program named BQCC822S merged the documented SRA data into the SPLIT2 group. It read the second BGP.BQC.DS722 tape and created BGP.BQC.DS822. Then another short program was run to swap out the SRA___ variables and replace the SQ___ and PAR___ variables. This new BGP.BQC.DS872 tape contained the best values and flag variables from the latest step in the merge, as well as

the SQ and PAR variables. Program BQCC672S then merged the noncertified parent and student values. One last variable swap brought the SRA variables back in and dropped PAR and SQ variables for good. The tape created in this program was BGP.BQC.DS672. The final merge program, BQCC872S, was then run to merge undocumented SRA values.

The SPLIT1 group, which contains the students with corrections to their status items, skips the SRA merging programs. Program BQCC672S was run on the first BGP.BQC.DS722 tape to merge in undocumented parent and student data. At this point we had reformatted SAS data tapes of the various interview data sources and tapes containing the best documented values from all the sources. Using programs BQCC631S and BQCC731S we computed frequencies on the errors made on BEOG applications versus the problems they reported with the applications.

The two best value tapes, BGP.BQC.DS872 with the SPLIT2 group and BGP.BQC.DS672 with the SPLIT1 group were then used as input into program BQCC020S. This program calculates student eligibility indexes [SEI] from best values and compares these with the SEIs from the SER. It creates variables with the discrepancy values between best value and reported items and variables to flag errors. The two BGP.BQC.DS020 tapes created were then concatenated into a single complete BGP.BQC.DS020.FINAL tape. The first two BGP.BQC.DS020 tapes were separately run through BQCC811S which calculates awards and award discrepancies for each student. The two BGP.BQC.DS811 tapes were then

concatenated into a single BGP.BQC.DS811.FINAL tape, as well. These FINAL tapes were used to compute the various frequency tables and percentages in Volume 2 of this report. Finally, the award calculation and discrepancy variables were merged with the computer applicant record data on tape BGP.BQC.SEIZERO by program BQCC030S.

CHAPTER 5

APPLICATION PROCESSOR DATA ENTRY ERROR ANALYSIS

As part of the study of the BEOG application processor, an analysis of error rate associated with data entry was conducted.

SAMPLE SELECTION

The applications of approximately one-quarter of the 4,000 BEOG recipients interviewed in the nationwide survey were selected for the analysis. In order to maximize the use of time and resources, only applications which originated from the three Multiple Data Entry [MDE] processors--the College Scholarship Service [CSS], the American College Testing [ACT] Program, and the Pennsylvania Higher Education Assistance Agency [PHEAA]--were chosen. (The MDE processors accounted for approximately 85 percent of all BEOG initial applications during the 1980-81 processing year.)

The sample was drawn using Westat's BEOG Master File of all interviewed students and the 1980-81 Central Processor's History/Correction File. Records from both files were matched on Social Security numbers [SSN] and the first two characters of the last name. From the extract of matched records, 500 CSS originated applications, 500 ACT originated applications, and 250 PHEAA originated applications were randomly selected.

ERROR TABULATION

The primary research question was "To what extent does what students write on their form actually become what is entered in

the BEOG application processing system?" To answer this, the original application forms of the sampled recipients were visually compared with the data contained on the processor's History/Correction File.

Each of the MDE processors was asked to provide photocopies of the original applications. These photocopies were placed in groups of 10 and attached to a corresponding computerized listing of data from the History/Correction File. Data were compared in these groups of 10. All discrepancies, apparent discrepancies, or oddities between the data on the application and the data on the computerized listing were carefully recorded. The following information about these discrepant cases was logged: (1) the MDE processor; (2) the recipient's SSN; (3) the data item(s) in error; (4) the item value on the original application and the corresponding value on the listing; (5) the apparent nature of the error; and (6) whether or not the entry error had been corrected in subsequent transactions. As a quality control check, 1 completed case in each group of 10 was reviewed by a supervisor. When an error that had gone unrecorded was discovered, all cases in the group were reviewed.

Once all cases had been compared, senior analysts reviewed the log of errors. If questions arose about the nature of a particular error, that case was retrieved and examined. The total number of errors was tabulated manually from the error log.