DOCUMENT RESUME

ED 271 061 HE 019 442

TITLE	Title IV Quality Control Policy Option: Final Report on the Evaluation of Data Element Reduction, Stage I, Phase V.
INSTITUTION	Advanced Technology, Inc., Reston, VA.
SPONS AGENCY	Department of Education, Washington, DC.
PUB DATE	Jun 85
NOTE	118p.; Legibility of tables may be marginal due to type quality.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Computation; Data Collection; Efficiency; *Eligibility; Error Patterns; *Evaluation Criteria; *Federal Aid; Financial Aid Applicants; Financial Policy; Higher Education; *Need Analysis (Student Financial Aid); Public Policy; Quality Control; *Student Financial Aid
IDENTIFIERS	*Department of Education; *Pell Grant Program

ABSTRACT

Efforts of the Department of Education (ED) to simplify the Pell Grant formula by reducing the number of data elements used to calculate awards (i.e., data element reduction) are evaluated. A framework is developed to assess the critical characteristics of individual data elements, to eliminate elements from the formula, and to develop proposals for data element reduction. Individual data items used in the Pell eligibility and award formulae are evaluated on the basis of five measures: budget impact, aggregate distributional impact, sensitivity, reliability, and verifiability. Included is a comparison of two simulations of a reduction in the number of data elements used in the Pell eligibility and award formulae. The two simulations, the standard and the error free simulations, are identical except for the data base used. Both simulations eliminate all but five data elements (adjusted gross income, federal taxes paid, nontaxable income, number in household, and number in postsecondary education). The applicant-based model and data base, and techniques used to adjust the ED applicant data base for the error patterns found in the Pell Stage III data, are described in appendixes. References and a substantial series of data tables are also appended. (SW)

TITLE IV QUALITY CONTROL POLICY OPTION:

FINAL REPORT ON THE EVALUATION OF DATA ELEMENT REDUCTION

STAGE I

PHASE V

JUNE 1985

Submitted By

ADVANCED TECHNOLOGY, INC.
12001 Sunrise Valley Drive
Reston, Virginia 22091

TABLE OF CONTENTS

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

1
INTRODUCTION

The Department of Education (ED) has been interested in simplifying the Pell Grant formula by reducing the number of data elements used to calculate awards. This endeavor is commonly called data element reduction. Three overarching objectives motivate ED's approach to data element reduction. A reduced Pell formula must:

- Maintain or enhance the ability of the program to efficiently identify the target population,
- Simplify, streamline and make more understandable the determination of program eligibility and resulting awards, and
- Reduce the program distortions associated with error-prone, difficult to verify data elements.

Any data element proposal is also subject to the following constraints:

- Minimize the redistributional effects caused by data element reduction, and
- Neutralize the potential budgetary impact.

These objectives are not easily achieved. In fact, past attempts to eliminate data elements from the Pell formula have faltered because policymakers have been unable to demonstrate that these objectives could be achieved subject to the constraints identified.

Past analyses of reduced Pell formulae have assumed that eliminating infrequently reported data elements to increase efficiency automatically decreased equity by adversely affecting the awards of groups of recipients (e.g. those with high medical/dental expenses). The current analysis suggests that data elements placed in the Pell formula to enhance equity may actually undermine equity by introducing reporting error that distorts award patterns. These data elements may not have their intended effects on targeted recipients and their elimination may actually increase equity. Thus, a reduced Pell formula could achieve both efficiency and equity without massive distortions to awards for the vast majority of recipients.

The current analysis of data element reduction uses an approach that is fundamentally different from past analyses. A framework was developed to assess the critical characteristics of individual data elements and rank them under known assumptions. The framework allows one to select elements to eliminate from the formula and, thus, alternative data element reduction proposals can be developed for analysis and comparison. One recent proposal for a five element formula is discussed in-depth in Chapter 3 of this report.

SUMMARY OF FINDINGS

The analysis has produced many useful findings concerning data element reduction, including:

- The analytic framework used in this analysis can be a powerful tool for developing rational, defensible data element reduction proposals.
- Pell Grant data elements can be ranked in an objective, value-free manner according to their impact on the program.
- Data elements can be identified for retention in the formula or elimination on the basis of this ranking.
- The analysis of the five data element Pell formula with a standard and an "error free" data base suggests almost identical patterns in individual awards:
  - few recipients lose large amounts (over $400-$500)
  - the neediest students, those receiving the highest awards, continue to receive high awards (98 percent receive within $200 of the maximum award)
  - a disproportionate number of recipients who lose eligibility received low awards ($500 or less under the full formula)
- The cost estimates using ED's standard data base, which contains reporting errors, must stand as official estimates of the likely cost of data element reduction. However, a comparison of the cost estimates produced by the standard and error free simulations provides a potential budget range for a five element formula ($2.6 billion using standard data, $2.4 billion using error free data).
- The analysis suggests that increased costs incurred by reducing the formula to five elements could be potentially "financed" simply by eliminating error from the remaining elements, rather than adjusting formula taxation rates upward.

More specific findings from both the assessment of individual data elements and the analysis of a five element Pell formula follow.

Assessment of Individual Elements

The assessment of the impact of individual data elements has demonstrated that this analytic framework is both an appropriate and effective policy tool. The framework has provided a means for systematically evaluating and ranking 17 data elements in the Pell eligibility and award formulae across five measures. The framework provides a means of integrating these discrete measures (budgetary and distributional impact, sensitivity, reliability and verifiability).

We have provided two examples of how such an integration can be conducted and demonstrated how the results of these examples can inform policymakers in their consideration of data element reduction. In the first example, using equal weights for all measures, we ranked the data elements and classified them into three groups: high (high rankings on most measures), moderate (mixed rankings on these measures), and low (low rankings on most measures).

The data items were classified in the example as follows:

High

- Adjusted Gross Income
- Social Security Education Benefits
- U.S. Taxes Paid
- Family Size Offset
- Employment Expense Offset

Moderate

- Net Home Equity
- Number in College
- Nontaxable Income
- Veteran's Education Benefits
- Elementary and Secondary Tuition

Low

- Dependent Student's Net Assets
- Net Investment Equity
- Dependent Student's Income
- Net Business/Farm Equity
- Student Marital Status
- Cash/Savings/Checking
- Unusual Medical/Dental Expenses

The example generally suggests that the data items in the low classification could be considered for elimination from the Pell formulae with minimum impact across the five measures (budgetary and distributional impacts, sensitivity, reliability, and verifiability). Those classified as moderate would require closer scrutiny and would have higher impact. Those classified as high, for all practical purposes, could not be eliminated without substantial impact to the program. An example using differential weights for the measures resulted in two changes in the rankings and no changes to the classifications.

The discussion above is only a summary of the examples. The results of these must be put into the context provided by the thorough discussion of the analysis, findings, and the caveats provided in Chapter 2.

Analysis of a Five Element Formula

As Chapter 2 presents a methodology and data for developing data element reduction proposals, Chapter 3 presents a detailed and thorough analysis of the budgetary and distributional impact of one data element reduction proposal, a five element formula. Two simulations, conducted for Advanced Technology by the Division of Policy and Program Development (DPPD), Office of Student Financial Assistance, formed the basis of the analysis. The first simulation used a standard applicant data base in conducting model runs of full and five element formulae. The second simulation was identical to the first except that an "error free" data base was used to simulate the effect of eliminating error along with data elements. (A description of the imputation procedures used to develop this unique data base is contained in Technical Appendix B.) A comparison of the two simulations produced the following findings:

- Differences in impact are most evident on the aggregate level of program costs and number of recipients.

- The error free simulation results in nearly 150,000 fewer recipients and a slightly higher budgetary impact than the standard simulation. However, the baseline budget was substantially lower (about $200 million) for the error free simulation.

- The error free simulation produces a lower baseline budget (about $2.2 billion) and the five element formula without taxation increases roughly equals the standard simulation full formula baseline costs (about $2.4 billion). This calls into question the need to increase taxation rates in the simulation.

- Average awards for the error free simulation are unchanged but lower than the standard simulation, in which awards decline.

- On most other dimensions (e.g., numbers of awards increasing, decreasing, or staying the same by applicant characteristic) the differences are minimal.

These findings and the analysis of the simulation are discussed in detail in Chapter 3.

BACKGROUND

Discussions surrounding the number and type of data elements used in determining eligibility and award for the Pell Grant program are as long-standing as the program itself. These discussions typically have focused on several major policy-relevant issues including the program costs for different combinations of data elements, the sensitivity of different formulae to specific groups of applicants, and the redistributive effects of adding or eliminating data elements. In addition, the relationship of the Pell formula to the overall student aid delivery system has been a concomitant issue.

Recently, the findings of the Pell Grant Quality Control (QC) Project have resurfaced data element reduction as a potential corrective action which could lower program-wide error through eliminating error-prone data elements from the Pell SAI and award formulae, and simplify the application process as well. The Pell Grant QC

Project measured quality in the delivery of funds in the Pell Grant Program. Using a variety of data collection methods, including institutional site visits, record abstractions, personal interviews with parents and students, and acquisition of IRS records, the project recomputed awards based on the most reliable data and then with original awards and institutional disbursements. The results of the project were twofold. First, the analyses generated program-wide estimates of errors; second, these analyses identified data elements in the SAI and award formulae that were error-prone and difficult to validate. Consequently, as part of the Title IV Quality Control Project, the Division of Quality Assurance (DQA) has identified Pell Grant data element reduction as a potential corrective action to reduce errors and has requested a series of analyses to support ED policymakers in the renewed policy discussion surrounding data element reduction.

Numerous analyses of data element reduction have been undertaken in recent years. Most have focused on the budgetary impact of reduction and the alteration of the award patterns that exist under the current formula, which are most often used as a measure of program equity. However, none of these analyses was able to analyze fully the impact of data element reduction for at least two reasons. First, most previous analyses assumed that reported application data were correct and hence failed to capture the effects of error on the program. Second, none of these recent analyses was able to systematically evaluate the impact of data elements across several diverse program goals.

Program-wide analyses of several combinations of data elements in a reduced eligibility formula conducted by Advanced Technology during Stage II of the Pell QC Project accounted for error by using verified data in the simulations.[1] Despite controlling for applicant error for the first time, these analyses were conducted on a recipient data base and therefore the impacts of these alternative combinations on newly eligible recipients could only be estimated. As a part of the present policy option, preliminary analyses were conducted to measure the program-wide effects of data element reduction at a detailed level.[2] These analyses utilized data from the official ED applicant-based model, with the assistance of the Pell Grant Branch,

[1] Compilation of Quality Control Findings: Information on Policy Options, March 1983.

[2] Title IV Quality Control Policy Option: Preliminary Analysis of a Simulated Five Data Element Pell Grant Eligibility Formula, September 1984.

DPPD, to measure the effects of data element reduction on subpopulations of applicants. While these data brought the strengths of an applicant data base to the analyses, the analyses could not account for application error, a major source of program error. However, the findings from the 1982-83 Pell Grant QC Project allow substitution of more accurate data for error-prone data elements through the creation of an adjusted applicant data base and measurement of the effects of data element reduction on the pattern of awards. This provides a more accurate basis for comparing distributions of awards under the full and reduced data element formulae. Both the preliminary and the present analyses of full and reduced formulae hold the budget constant by adjusting upward the taxation rates.

Another approach to data element reduction was proposed by Advanced Technology. An informal position paper presented a framework for systematically evaluating the impact of individual data elements. The Stage III Corrective Actions volume[3] from the Pell QC Project utilized this framework and presented an approximation of the impact of each element across five criteria, using Stage III Pell recipient data.

This policy option report represents an integration of the approaches from several prior analyses and benefits from the strengths of each. The analysis has two discrete parts. The first, which was recommended in the Stage III Corrective Actions volume, assesses the impact of individual data elements on five program dimensions:

- Budgetary Impact
- Aggregate Distributional Impact
- Sensitivity
- Reliability
- Verifiability

---

[3] Quality in the Pell Grant Delivery System, Volume 2, Corrective Actions, April 1984, pp. 4-8 through 4-13.

These dimensions and the assessment methodology are described in Chapter 2 of this report.

The second analysis compares distributional trends resulting from program-wide simulations of the applicant-based model for the full formula used for the 1982-83 academic year with a five element formula using both reported data (those containing error) and error adjusted or "best" data (from which error found in the Pell QC Stage III has been corrected). Chapter 3 contains this analysis.[4]

ANALYTIC CONTEXT

The nature and focus of the analysis conducted for this policy option report must be carefully delineated and explicitly contrasted with policymaking. Both analyses--the program-wide simulation of the full and five element formulae and the assessment of the impact of individual data elements--have been designed to provide data with which ED policymakers can make informed policy decisions. We have avoided making implicit policy decisions throughout our analysis. For example, the goal of assessing individual data elements is to provide policymakers with a framework for ranking data elements according to their impact, not to advance any one proposal within this paper. Nevertheless, analysis such as this requires making judgments in order to provide data to ED for policymaking purposes. We have clearly identified points at which judgments were made and explicitly stated these judgments.

In addition, the policy relevance of the findings must be delineated carefully, particularly with regard to simulating the program-wide effects of reducing the number of data elements in the Pell eligibility and award formula to five. The analysis has been designed as an **evaluation**, not as a forecast. The emphasis of the assessment of individual data elements is the measurement of the impact of data elements across several dimensions. Therefore, the findings from both analyses can isolate the effects of data element reduction within a research context; only official ED estimates can stand as forecasts of likely policy consequences.

Some general comments should be offered concerning the data base, simulations and generalizability of the results of our analyses. These simulations utilize a large

[4] Technical Appendix A contains descriptions of the ED model, applicant data base, and the full and five element formulae simulations.

data base that permits generalization to the population of applicants. Different eligibility criteria, however, are likely to change the composition of the applicant population. We were unable to account for this likelihood in this analysis, since the model and our analyses simulate the effects of program changes on an existing and static applicant population. Also, the results of the assessment of individual data elements are, to a degree, formula specific, although some of the results would be identical. The degree of difference between the formula used and another--a subsequent year or reduced form--must be examined and considered before generalizations could be considered. This analysis focuses explicitly on the impact of eliminating data elements from the eligibility and award formulae. It does not assess the implications of eliminating items from the application form nor does it deal with issues of compatibility with other need analysis tests or forms. Although these are important considerations, they are beyond the scope of this analysis.

This analysis can play the important role of informing the policy debate by measuring the efficiency of data element reduction as a corrective action for program error by accurately and comprehensively capturing its effects. The assessment of individual data elements can also serve as a basis for developing alternative proposals for altering the number and types of data elements used in the determination of eligibility and award.

ORGANIZATION OF THE REPORT

This report is comprised of two chapters that parallel the analysis and technical appendices. Chapter 2 describes the analysis and findings resulting from the evaluation of the marginal impact of the individual data elements. Chapter 3 compares two simulations of a reduction in the number of data elements used in the Pell eligibility and award formulae using two data bases. The Appendices describe the data base and model, the imputation that was conducted to adjust the ED applicant data base for the error patterns found in the Pell Stage III data, and additional program simulation tables.

2

EVALUATION OF THE IMPACT OF INDIVIDUAL DATA ITEMS

Characteristically, data element reduction has been approached by presenting alternative configurations of eligibility formulae with five, six, or seven data elements or substituting number of exemptions for household size. These alternatives have then been evaluated by measuring changes to the budget and the distribution of awards at the program level induced by changing the formula. Despite the intuitive appeal and relative ease of such an approach, these analyses have failed to provide either a framework or the data for systematically developing and evaluating alternatives. In addition, the development and evaluation of data element reduction alternatives are subject to competing, if not conflicting, goals which most approaches cannot deal with easily.

Data element reduction most often has been advanced as a strategy to maximize two of these program goals: integrity and efficiency. Integrity is maximized by making the program less error prone and increasing the reliability of data collected. Efficiency is achieved by reducing applicant data burden, administrative costs to institutions and application processing costs to the government. However, past reduction proposals have run afoul of budget and equity concerns. Analyses of data element reduction proposals have suggested that these proposals cause budget increases and shifts in distribution of awards that were judged to be unacceptable and resulted in decreased program sensitivity to applicant characteristics. Prior policy discussions have not provided the framework or data with which to consider these goals simultaneously.

The current approach provides both the framework and the data with which to make informed judgments about alternative configurations of data elements. This approach provides these by evaluating each data element individually on the basis of five measures:

- Budget Impact
- Aggregate Distributional Impact
- Sensitivity
- Reliability
- Verifiability

The approach also ranks the data elements for each measure ordinally from the highest to the lowest impact.

This approach also allows for simultaneous consideration of these measures in order to enable policymakers to identify groups of items that must remain in the formulae, those that can be eliminated with little impact, and those that could be eliminated given certain tradeoffs. An underlying premise of the analysis suggests that items that rank low on all measures more easily could be eliminated, whereas high-ranking items should be retained.

METHODOLOGY AND ANALYSIS

The focus of the analysis in this portion of the report is the evaluation of data items used in the eligibility and award formulae as they directly affect the award. For the most part, these data elements correspond with a single formula item.[5]

Each item was evaluated individually changing to zero all non-zero reported data values for the item being evaluated, such as net home value or unusual medical and dental expenses. Table 1 lists the values used to eliminate the item from the formula. All awards were then recalculated and analyzed for each of the five measures. For one item, family size offset, changes to the SAI software were necessary in order to eliminate the data item.

Measures and Database

In this portion of the analysis five measures are used to assess the impact of individual data elements on awards. In order to assess this impact we used the 1982-83

[5] Two exceptions are Family Size and Marital Status which affect multiple formula elements.

TABLE 1

DATA ITEMS EVALUATED THROUGH ELIMINATION FROM THE PELL ELIGIBILITY AND AWARD FORMULAE

Data Item	Value Used to Eliminate the Data Item
Income	
Adjusted Gross Income	0
Nontaxable Income	0
U.S. Taxes Paid	0
Dependent Student's Income	0
Veteran's Education Benefits	0
Social Security Education Benefits	0
Assets	
Net Home Equity	0
Net Investment Equity	0
Cash/Savings/Checking	0
Net Business/Farm Equity	0
Dependent Student's Net Assets	0
Offsets and Protections	
Student's Marital Status	Unmarried
Family Size Offset	0
Number in College	1
Unusual Medical and Dental Expenses	0
Elementary and Secondary Tuition and Fees	0
Employment Expense Offset	0

ED data base and a standard full formula for the 1982-83 program year as a baseline. Individual data elements were removed from the formula and awards were recomputed using the 1982-83 Pell eligibility and award formulae. The resulting awards were multiplied by a sampling weight assigned to each applicant on the file and, from the first two measures, by a participation rate assigned by income level. These procedures estimate program changes attributable to the elimination of the data element. The changes were then analyzed through the five measures, each of which is described below.

- **Budgetary Impact** is the change in program budget when a data element is excluded and the resulting budget is compared with the baseline budget under a full formula.
- **Aggregate Distributional Impact** is measured as the change in the distribution of program funds across income and other categories compared against the baseline distribution with all elements included in the formula.
- **Sensitivity** is a measure of the relative responsiveness of the program to applicants with particular characteristics (e.g., two working parents). Sensitivity is reported as the average change between the base award and the recomputed award with the data item removed.
- **Reliability** is the degree to which reported data accurately represent applicants' true characteristics.
- **Verifiability** is an assessment of the degree to which items can be checked against reliable corroborative data sources.

The framework utilized requires that we make judgments concerning several analytic issues including classification and weighting. In each of the analyses, data elements are classified as having high, moderate, or low impact. The basis upon which data elements were assigned to these categories is explicitly treated in each of the following sections. In the last section of this chapter, the results of the five analyses are integrated. Although we have included two examples of weighting schemes, the values we assigned to the classifications in order to rank the data items (2, 1, 0 for high, medium, low) remain constant. The use of different values (for example, 5, 1, 0, respectively) may alter the ranking and potentially the classification.

The remainder of this chapter is divided into sections that describe the analysis conducted for each of these measures and the findings of these analyses. Each measure addresses a specific research question that introduces the sections.

Budgetary Impact

One of the primary and often asked questions concerning the effects of data element reduction is the impact on the program budget. This portion of our analysis was motivated by the following question: **How does the program budget change when single data elements are removed from the Pell formulae?** Within this framework, data elements that had high budgetary impact would likely be retained in the formula; those with low budgetary impact would be candidates for elimination on the basis of budgetary impact.

In order to address this question, we eliminated each of the 17 data items in turn and recomputed awards for cases in which changes to the data element were made and summed all weighted awards. The result was a new program budget total. The difference between the baseline budget and the new budget is defined as the budgetary impact, represented as a dollar difference and percentage change. Table 2 represents the ranking of the budgetary impact of removing individual data elements. The data elements are ranked from highest to lowest percent absolute change. In addition these budgetary changes are classified as high, moderate or low according to the following ranges:

- High -- more than 10 percent change in program cost (approximately $250 million)
- Moderate -- 2 to 10 percent change in program costs (approximately $50 to $250 million)
- Low -- less than 2 percent change in program costs (approximately $50 million or less)

Several features of Table 2 are noteworthy. Eliminating data elements produces both positive and negative changes. Increases in budget result from eliminating income or asset items that are used as resources for family contribution to educational costs. Conversely, decreases in budget result from eliminating expense allowances that protect portions of income from contribution. Adjusted gross income, family size, and social security education have the greatest budgetary impact, although the changes are both positive and negative. Adjusted gross income, family size, and social security education benefits have the greatest budgetary impact, although the changes are both positive and negative. Seven data items (VA education benefits, elementary and secondary tuition, investment equity, business farm equity, cash/savings, student's marital status and medical/dental expenses) affect program

TABLE 2

RANKING OF THE BUDGETARY IMPACT OF ELIMINATING DATA ELEMENTS FROM THE ELIGIBILITY AND AWARD FORMULAE

Impact	Data Item	Increase/ Decrease(-) in Program Budget[1] $ (millions)	Percent Change in Program Budget
HIGH	Adjusted Gross Income	1708	68.66
	Family Size Offset	-1455	-58.49
	Social Security Education Benefits[2]	276	11.10
MODERATE	U.S. Taxes Paid	-155	-6.23
	Net Home Equity	117	4.72
	Number in College	-110	-4.42
	Nontaxable Income	90	3.64
	Employment Expense Offset	-80	-3.23
	Dependent Student's Income	71	2.86
LOW	Dependent Student's Net Assets	35	1.39
	Veteran's Education Benefits	13	0.53
	Elementary and Secondary Tuition	-13	-0.53
	Net Investment Equity	10	0.39
	Net Business/Farm Equity	8	0.34[3]
	Cash/Savings/Checking	8	0.30[3]
	Student's Marital Status	5	0.21
	Unusual Medical/Dental Expenses	-2	-0.08

[1] Baseline Budget is $2,488 million.

[2] The Pell formula no longer contains social security education benefits. It is not possible in this analysis to estimate with any accuracy the impact of eliminating this data element from different formulae. However, the effects are not likely to challenge the findings of this analysis.

[3] Difference due to rounding.

costs by less than 1 percent. Several of these items, the asset items, are subject to $25,000 protections and are, for most applicants, "taxed" at 5 percent, effectively reducing the budgetary impact of these items. Relatively few applicants report tuition expenses or levels of medical expenses high enough (greater than 20 percent of effective family income) to reduce family discretionary income.

This analysis uncovers an interesting, seemingly anomalous, finding relating to the difference between the impact of social security and veteran's education benefits. Both of these elements are included in the award formula, which means that they more directly affect Pell awards than other elements in the SAI formula that are taxed or subject to protections or offsets. However, the budgetary impact of VA education benefits is vastly lower than social security education benefits. This is a result of the fact that far fewer (about 2 percent) report receiving VA benefits as opposed to social security (about 11 percent). The mean value for VA benefits ($3,200) is also slightly more than half the mean value for social security ($5,300). These two facts result in a substantially lower budgetary impact for VA benefits. This, of course, is to be expected. Items that were infrequently reported or had low effective values tended to have low budgetary impact.

Aggregate Distributional Impact

The impact on the distribution of awards resulting from changes to the eligibility and award formulae is of fundamental importance to any analysis on the impact of data elements. Particularly since the impetus for data element reduction is the reduction of error, rather than redirecting program funds, the elimination of data elements from the formulae must have as a constraint minimizing redistributive effects induced by these changes. Therefore, a particularly relevant question for this analysis is: **What is the impact on the distribution of awards of eliminating each of the 17 data elements?** Data elements that have high redistributional impact on program funds would likely be retained; those that have low redistributional effects would be candidates for elimination.

This distributional analysis was conducted by comparing the applicant's original award under the full formula with the award when the respective data element was removed from the formula. The results of these comparisons, for presentational purposes, were tabulated by percentage of applicants who experienced no change in award (+/-$100) and two levels of increases and decreases ($101-$600 and over $600) and ranked from highest to lowest impact. Those data items that induced the largest number of increased and/or decreased awards were ranked as having the highest distributional impact. Conversely, the data items that cause the fewest changes in awards were ranked as low impact.

Table 3 presents the results of this distributional analysis and an ordinal ranking of the distributional impact of each individual data element. In addition, the distributional effects are classified as high, moderate, or low in the following manner:

- High -- Greater than 10 percent of the applicants would receive a different award (different by more than $100) when compared with the original award.
- Moderate -- Greater than 5 percent but less than 10 percent of the applicants would receive a different award (different by more than $100) when compared with the original award.
- Low -- Less than 5 percent of the applicants would receive a different award (different by more than $100) when compared with the original award.

Several conclusions can be drawn from the table about the distributional impact of individual data elements. Only three data elements cause redistribution for more than ten percent of all applicants (family size, adjusted gross income and U.S. taxes paid) and therefore could be considered to have high impact. Four more data elements can be classified as having moderate impact, causing redistribution in between five and ten percent. Ten data elements have a redistributive impact for less than five percent and are considered to have low impact. Six of these 10 low impact data elements cause redistribution for less than one percent of all applicants.

Sensitivity

The preceeding measures assess the impact of eliminating data elements at a program-wide or aggregate level. Although this assessment is fundamental to any analysis of changes to the Pell formulae, other dimensions of the impact cannot be overlooked, including the effects of the change in awards of individual applicants.

TABLE 3

RANKING OF THE IMPACT ON DISTRIBUTION OF AWARDS OF ELIMINATING INDIVIDUAL DATA ELEMENTS FROM THE ELIGIBILITY AND AWARD FORMULAE RANKED FROM HIGHEST TO LOWEST IMPACT

Impact	Data Element Eliminated	Increase: Over $600 (%)	Increase: $101 to $600 (%)	No Change (+/- 100) (%)	Decrease: $101 to $600 (%)	Decrease: Over $600 (%)
HIGH	Family Size Offset	0	0	49.41	19.89	30.70
	Adjusted Gross Income	32.26	9.86	57.87	0	0
	U.S. Taxes Paid	0	0	85.19	14.66	0.15
MODERATE	Employment Expense Offset*	0	0	91.11	8.82	0.07
	Number in College	0	0	91.90	7.17	0.93
	Social Security Education Benefits	6.08	1.93	91.99	0	0
	Net Home Equity	1.82	4.70	93.47	0	0
LOW	Nontaxable Income	1.40	3.14	95.46	0	0
	Dependent Student's Income	1.38	1.59	97.03	0	0
	Dependent Student's Net Assets	0.24	2.57	97.20	0	0
	Elementary and Secondary Tuition	0	0	98.76	1.20	0.04
	Veteran's Education Benefits	0.27	0.47	99.27	0	0
	Student's Marital Status	0	0.63	99.29	0.04	0.04
	Cash/Savings/Checking	0.07	0.50	99.43	0	0
	Net Real Estate/Investment Equity	0.16	0.32	99.53	0	0
	Net Business/Farm Equity	0.16	0.11	99.72	0	0
	Unusual Medical and Dental Expenses	0	0	99.88	0.09	0.02

*Not an application item, computed from income portions.

Therefore, this analysis explored another research question: How do individual awards change for applicants facing particular circumstances when a data element, included in the formula to sensitize the award to such circumstances, is removed?

Elimination of data elements from the formulae can have a substantial effect on the sensitivity of the formulae to specific groups of applicants, an important component of equity. Equity, as it is used in this context, can be stated simply as equal treatment of equals. The Pell formulae (eligibility and award) have many components that potentially enhance sensitivity--the ability to account for differences among applicants--and thereby equity.

Elimination of data elements can decrease sensitivity by reducing the ability to differentiate among applicants. In addition, elimination of certain data elements will affect sensitivity to a greater degree than others. For example, the elimination of the family size offset would certainly have a greater impact on sensitivity than the elimination of medical/dental expenses, since the former decreases discretionary income by approximately $1,200 for each additional family member from a base of $4,200 and the latter reduces discretionary income by the amount of expenses in excess of 20 percent of effective income (all income minus taxes). Those data elements that are included in the formula to enhance sensitivity but have little impact on awards--even for applicants at the upper ranges of the data value--would be candidates for elimination on the basis of sensitivity.

We have measured the impact on sensitivity of awards to the individual data elements by identifying the upper range of data values,[6] eliminating the value and recomputing the award for this subsample of cases. Table 4 lists the data values for these ranges. The upper range of each value was selected because the elimination of the data elements would show the greatest impact at that level.

---

[6]The range selected for most data elements was the 90th and 95th percentile. This measures the maximum impact of the data element on the award while avoiding biasing the measure by including outliers. For several data items (elementary and secondary tuition, net business/farm equity, net investment equity and veteran's educational benefits) the values between the 90th and 95th percentile were zero, consequently we measured award changes for values between the 95th and 99th percentile.

TABLE 4

VALUES FOR DATA ELEMENTS USED IN THE SENSITIVITY ANALYSIS

	Range of Data Values[1] Low	High
Adjusted Gross Income	29,084	31,464
Social Security Education Benefits	1,005	4,963
Net Home Equity	38,220	49,879
U.S. Taxes Paid	4,418	5,351
Family Size	6	7
Employment Expense Offset	1,500	1,500
Number in College	2	4
Nontaxable Income	5,078	7,932
Veteran's Eduction Benefits[2]	1	4,699
Elementary and Secondary Tuition[2]	563	2,052
Dependent Student's Net Assets	159	533
Net Investment Equity[2]	6,882	40,145
Dependent Student's Income	2,387	3,694
Student Marital Status	married	married
Cash/Savings/Checking	3,001	6,103
Unusual Medical/Dental Expenses	1,139	1,629
Business/Farm Equity[2]	4,180	77,730

[1] All values are in the 90th to 95th percentile range unless otherwise noted.

[2] These values are in the 95th to 99th percentile range because the value of the 90th percentile was zero.

It should be noted that we measured sensitivity for all data elements with the single exception of dependency status, which posed methodological problems. Clearly, the elimination of several of these, such as AGI, would not seriously be considered, since this would alter the fundamental nature of Pell as a need-based student aid program. Nevertheless, these elements were included in the analysis in order that the methodology be comprehensive, and the ranking of the elements be accurate.

Table 5 presents the results of this analysis of sensitivity. The table ranks the data elements on the basis of absolute percent change in award. In addition, the sensitivity of the data element is classified as high, moderate, or low in to the following manner:

- High -- 50 percent or greater change in mean award
- Moderate -- 10 percent or greater but less than 50 percent change in mean award
- Low -- 10 percent or less change in mean award.

Table 5 contains several columns: the base or original award, the marginal award recomputed with the respective data element eliminated, the change in award or difference between the two, and percent change in award. The change in award represents the sensitivity of the award to the data element measured in dollars. The percent change in award represents the change in award as a percentage of the mean baseline award. The data items are ranked on the basis of absolute percentage change in award from highest (AGI, 1,507 percent) to lowest (business/farm equity, .1 percent), ignoring the direction of the change. Items were ranked by absolute change because it was assumed that increases and decreases have equal weight; that one is not preferential to the other from the perspective of sensitivity. The data in Table 5 suggest that, given the methodology, awards are most sensitive to the high impact elements, including AGI, social security education benefits, net home equity, U.S. taxes paid, and family size. The relatively low mean baseline award for AGI ($81) results from the fact that few applicants with AGI's within the 90 to 95th percentile receive awards. Thus, the mean or average award is depressed by the large number of zero awards in that range of AGI values. When AGI is eliminated from the formula, awards increase dramatically, because of the nature of the formula. Awards have relatively high sensitivity to social security education benefits because these benefits directly reduce award since it is part of the award formula.

TABLE 5

SENSITIVITY OF AWARD TO THE ELIMINATION OF INDIVIDUAL DATA ELEMENTS BY DATA ELEMENT

		Mean		Difference	
Sensitivity	**Data Item**	**Base Award**[1]	**Marginal Award**[2]	**Change In Award**	**Percent Change In Award**
HIGH	Adjusted Gross Income	81.54	1,310.4	1,228.86	1,507.06
	Social Security Education Benefits	315.52	928.54	613.02	194.29
	Net Home Equity	171.84	344.09	172.25	100.24
	U.S. Taxes Paid	58.45	8.85	-49.6	-84.86
	Family Size Offset	606.15	213.58	-392.57	-64.76
MODERATE	Employment Expense Offset	89.74	65.41	-24.33	-27.11
	Number in College	579.27	478.12	-101.15	-17.46
	Nontaxable Income	569.52	647.70	78.18	13.73
	Veteran's Education Benefits	676.96	760.08	83.12	12.28
	Elementary and Secondary Tuition	452.36	403.81	-48.55	-10.73
	Dependent Student's Net Assets	323.01	356.58	33.57	10.39
LOW	Net Investment Equity	270.97	292.31	21.34	7.88
	Dependent Student's Income	401.06	425.69	24.63	6.14
	Student's Marital Status	755.95	769.03	13.08	1.73
	Cash/Savings/Checking	267.89	271.36	3.47	1.3
	Unusual Medical/Dental Expenses	335.53	334.90	-.63	-.19
	Net Business/Farm Equity	603.24	603.82	.58	.1

[1]Original award computed with all data elements.

[2]Award computed with the respective data element eliminated.

Awards are moderately sensitive to six data elements ranging from employment expense offset (-27 percent) to dependent student's net assets (10 percent). Awards are relatively insensitive to another six elements. These range from net investment equity (8 percent) to business/farm equity (less than 1 percent).

Reliability

Program integrity is a fundamental design and program goal. In fact, if the data collected are not accurate and reliable, other program goals are undermined. Consequently, the reliability of applicant data is a relevant, if not essential, component of any assessment of the impact of individual data elements. We addressed this dimension of the analysis by posing the question: **How accurately does applicant reported data represent an applicant's true characteristics?**

The reliability of data elements was assessed through the use of the Pell Grant Quality Control Project Stage III data. We have defined reliability as the discrepancy rate found in Stage III. Two error rates were developed in this study: simple case discrepancy and case discrepancy with payment consequences. Case discrepancy occurs when true or validated data differ from application data used in the determination of Pell eligibility and award. Case discrepancy leads to payment consequences when the validated data result in a different award than calculated with original application data. Table 6 presents the discrepancy rates under both definitions and the ordered ranking for both. The data elements are ordered by case discrepancy rate. This rate was selected because it is more reliable since the other rate is formula specific and would change under a different formula. Thus, the former is more generalizable.

Data elements are also classified into groups of high, moderate, and low reliability items. This classification is the obverse of the error rate: the lower the error rate, the higher the reliability. The classification is as follows:

- <u>High</u> -- Less than 5 percent cases discrepant
- <u>Moderate</u> -- 5 to 10 percent cases discrepant
- <u>Low</u> -- Greater than 10 percent cases discrepant

TABLE 6

RELIABILITY OF DATA ELEMENTS USED IN THE
PELL GRANT FORMULAE RANKED FROM MOST TO LEAST RELIABLE

Reliability	Data Items	Cases with Discrepancies (%)	Rank	Cases with Discrepancies Resulting in Payment Consequences (%)	Rank
HIGH	Business/Farm Equity[1]	1.0	1	.1	1
	Veteran's Education Benefits	1.4	2	.6	3
	Net Investment Equity[1]	2.1	3	.3	2
	Elementary and Secondary Tuition	2.3	4	.7	4
MODER-ATE	Social Security Education Benefits	5.2	5	2.6	9
	Student's Marital Status	9.9	6	3.2	10
LOW	Net Home Equity[1]	10.7	7	1.8	13
	U.S. Taxes Paid	14.1	8	3.5	10
	Number in College	14.3	9	5.9	12
	Adjusted Gross Income	16.4	10	4.1	11
	Employment Expense Offset[2]	17.7	11	1.5[2]	7
	Family Size Offset	22.4	12	10.1	15
	Unusual Medical/Dental Expenses	23.2	13	.9	6
	Nontaxable Income	30.6	14	10.0	14
	Dependent Student's Assets	35.1	15	18.1	17
	Dependent Student's Income	37.0	16	14.5	16
	Cash/Savings/Checking	46.4	17	.8	5

[1]Estimate, computed from error rates for assets and debts.

[2]Estimate, computed from the error rate for income portions.

Four items in Table 6 have high reliability and their discrepancy rates and rankings are similar. Two are moderately reliable, although the rankings begin to diverge slightly for these items. Eleven items are classified as having low reliability based on case discrepancy rate. These range from net home equity (about 11 percent) to the least reliable, on this scale, cash/savings/checking (about 46 percent). Four items have low reliability using both rates: dependent student assets and income, nontaxable income and family size.

The rates differ because of the nature of the formula. Clearly, the more directly a change in the data element produces a change in award, the closer the rates and ranking. Many elements, such as cash/savings/checking, dependent student's assets and income, are subject to protections and taxed at a low rate; thus, the differences between the rates and rankings are wider.

Several observations should be made concerning this data and case discrepancy rate. First, the data are recipient data. We are consciously generalizing from recipient to applicant behavior. We believe this is sound because no data suggest that applicant and recipient misreporting behavior is different. In fact, the Title IV Quality Control Project, which examined error in the Campus-Based and Guaranteed Student Loan Program and included many Pell applicant non-recipients, reports error patterns generally similar to the Pell QC Project. Second, the discrepancy rate represents the rate at which the true or validated data values differed from reported values by more than plus or minus $2, the range specified by ED in the Pell QC Project. Third, the rate includes zero and non-zero reported values. Since the discrepancy reflects both values, the rates are themselves an artifact of the occurrence of this characteristic in the general population. For instance, if a small percentage of the population has business/farm equity, the error rate inherently will be lower than for AGI or nontaxable income. This occurs because, among other reasons, nonbusiness/farm owner applicants implicitly report zero values. Thus, there is a lower probability of error in the general population.

Verifiability

The final dimension on which the data elements were evaluated is verifiability. Verifiability is a corollary of reliability and a logical and important policy concern in any systematic evaluation of data elements. We focused our analysis by addressing the question: **To what degree can the data element be corroborated through an alternate source of documentation?**

Our assessment of the degree to which data elements can be verified is essentially qualitative. The assessment draws upon a rich body of qualitative data developed through the fall 1982 study of Pell validation compliance and particularly the "best value" selection software for the Pell and Title IV QC projects. The research that produced the best value selection software and documentation represents one of the most thorough reviews of corroborative documentation for data items used in the Pell formula. These data informed our assessment of individual data items.

Each item was analyzed from five perspectives:

- Is a reliable corroborative data source available for each item?

  In answering this question, we essentially asked whether a document existed with which the data item could be verified and which was produced by an "official," neutral third party. We also considered whether the data from this document treated the time period and used the same general definition for the data item as the formula.

- Is the document readily available?

  In assessing the data element from this perspective we considered whether most families have and maintain this documentation. Conversely, if families must request the document often, we considered whether it was easily obtained. The experience of our staff's fieldwork with financial aid staff was used extensively in this analysis.

- Is the document provided quickly?

  Here we evaluated whether the agencies (companies, etc.) from which a family would have to request a document(s) provide these in a timely manner. We also called upon staff experience with financial aid officers, and their experiences, to conduct this evaluation.

- Is the data retrospective?

  We assessed whether the data used in the formula was retrospective (e.g., prior or base year AGI), which can be verified more easily.

- Can errors of omission as well as commission be detected?

  Lastly, we evaluated the degree to which failing to report as well as under or overreporting could be identified.

These five questions focused our assessment of the individual data elements. Once each data item was evaluated, we ordinally ranked the items. Ranking took place in several stages. Each of the questions discussed above was weighted equally, except omission/commission, which doubled the elements' score if both errors could be detected. Each of the data elements received one of five assessments (yes, reliable approximation, uncertain, often no, no). Each of these was weighted on a symmetrical scale from +2 for yes to -2 for no. The elements were then classified into high, medium, and low error of validation as follows:

- High -- Three or more yeses and both omission/commission (a score of greater than 10)
- Moderate -- Between two yeses and both omission/commission, and three yeses (a score of between 6 and 10)
- Low -- Fewer than three yeses (a score of less than 6)

Table 7 presents the results of the evaluation. Four elements are classified as having high verifiability; four as moderate. Nontaxable income is ranked by the composite of its subcomponents, which are examples of the types of income that are included in this data element.

The verifiability for the remaining data elements is classified as low. Generally these are asset items (home, business/farm, and investment equity and dependent student assets), demographic items (family size, number in college and student's marital status) and expenses (medical/dental). Assets receive low scores because of the difficulty of establishing value, the relative difficulty in discovering errors of omission and the potential difficulty of rapidly providing up-to-date documentation. Two of the demographic items, family size and number in college, are prospective and therefore virtually unverifiable, although number of exemptions can be used as a reasonable approximation, acknowledging the limitation of such comparisons. Student's marital status is difficult to verify because almost nothing short of a marriage license can conclusively prove the student's status. Therefore, no other documentation can be considered reliable (e.g., tax forms). Medical/dental expenses may be difficult to verify simply because of the potential volume and diversity of documentation and payment forms.

TABLE 7

VERIFIABILITY OF DATA ELEMENTS USED IN THE PELL FORMULAE

Classification	Item/Sub-Item	Reliable Source	Readily Available	Provided Quickly	Retrospective	Omission/ Commission
HIGH	Adjusted Gross Income	Yes	Yes	Yes	Yes	O/C
	Employment Expense Offset	Yes	Yes	Yes	Yes	O/C
	U.S. Taxes Paid	Yes	Yes	Yes	Yes	O/C
	Veteran's Education Benefits	Yes	Yes	Yes	No	O/C
MODERATE	Social Security Education Benefits	Yes	Yes	Yes	Yes	C
	Dependent Student's Income	Yes	Yes	Yes	Yes	C
	Cash/Savings/Checking	Reliable Approximation	Yes	Yes	Yes	C
	Nontaxable Income					
	Social Security Benefits	Yes	Yes	Uncertain	Yes	C
	AFDC	Yes	Uncertain	Often No	Yes	C
	Child Support	Often no	Often No	Uncertain	Yes	C
	Welfare	Yes	Uncertain	Uncertain	Yes	C
	Unemployment	Yes	Yes	Uncertain	Yes	O/C
	Railroad Retirement Benefits	Yes	Yes	Yes	Yes	C
	Disability Income	Yes	Yes	Yes	Yes	C
	Veteran's Benefits	Yes	Yes	Yes	Yes	C
	Interest from Tax Free Bonds	Yes	Yes	Uncertain	Yes	C
	Elementary/Secondary Tuition	Yes	Yes	Uncertain	Yes	C
LOW	Dependent Student's Net Assets	Reliable Approximation	Yes	Uncertain	Yes	C
	Net Home Equity	Reliable Approximation	No/Uncertain	Often No	Yes	C
	Net Investment Equity	Reliable Approximation	Uncertain	Often No	Yes	C
	Net Business/Farm Equity	Reliable Approximation	Uncertain	Often No	Yes	C
	Unusual Medical/Dental Expenses	No	No	No	Yes	C
	Student's Marital Status	No	No	No	Yes	
	Family Size Offset*	No	No	No	No	O/C
	Number in College*	No	No	No	No	C

*Prospective items; evaluation in future years.

Joint Consideration of the Measures

The analyses presented in the prior sections of the chapter provide the data with which to evaluate the impact of individual data elements across several measures. However, we have assumed that decisions concerning the elimination of data elements cannot be made on the basis of any single measure or dimension. Consequently, our approach has assumed that it is necessary to jointly consider the impact of data elements across these five dimensions. Such an integration, however, confronts fundamental policy questions, for instance concerning the relative importance of each of the measures, which only F.J policymakers can address. Fully acknowledging this fact and the fact that policymakers may differ concerning the relative importance, our approach to integrating the results of the discrete analyses is two-fold. First, we present a framework that allows ED policymakers to make individual judgments about the impact of data elements. Second, we provide two examples of how such judgments can be made within this framework.

There are numerous ways to classify the data elements across the five measures. For brevity's sake, we have chosen only two as examples. Ta`le 8 presents the first such example. In this first example we assume that each of the measures has equal importance and therefore high budgetary impact is equally as important as high reliability and verifiability. In addition, for simplicity's sake, we have grouped the data elements by assigning values to high, moderate, and low scores (2, 1 and 0, respectively) on each of the measures and divided the elemer ; into three approximately equal high, moderate, and low classes. Those elements classified as high on average have the highest impact across the five measures; conversely, those classified as low have the lowest. We have assumed that one would approach the elimination of data elements by beginning with data elements in the low joint classification and considering whether the elimination of each data elemei t requires too substantial a tradeoff.

One of the seven data elements in the low joint classification (medical/dental expenses) received low classification across all of the measures. Dependent student's income had moderate budgetary impact and verifiability. Dependent student's net

TABLE 8

EXAMPLE OF JOINT RANKING OF THE DATA ELEMENTS
ASSIGNING EQUAL WEIGHTS TO EACH MEASURE

Classification		Budgetary Impact (Weight=1) ($ Million)	Distributional Impact (Weight=1) (%Δ)	Sensitivity (Weight=1) (%Δ in award)	Reliability (Weight=1) (% w/error)	Verifiability (Weight=1) (Rank)
HIGH	Adjusted Gross Income	High (1,708)	High (42)	High (1,507)	Low (16)	High (1)
	Social Security Education Benefits	High (276)	Moderate (8)	High (194)	Moderate (5)	Moderate (5)
	U.S. Taxes Paid	Moderate (-155)	High (15)	High (-85)	Low (14)	High (3)
	Family Size Offset	High (-1,455)	High (51)	High (-65)	Low (22)	Low (16)
	Employment Expense Offset	Moderate (-80)	Moderate (10)	Moderate (-27)	Low (18)	High (2)
MODERATE	Net Home Equity	Moderate (117)	Moderate (7)	High (100)	Low (10.7)	Low (11)
	Number in College	Moderate (-100)	Moderate (8)	Moderate (-17)	Low (14)	Low (17)
	Nontaxable Income	Moderate (90)	Low (5)	Moderate (14)	Low (31)	Moderate (8)
	Veteran's Education Benefits	Low (13)	Low (1)	Moderate (12)	High (1)	Moderate (4)
	Elementary and Secondary Tuition	Low (-13)	Low (1)	Moderate (-11)	High (2)	Moderate (9)
LOW	Dependent Student's Net Assets	Low (35)	Low (3)	Moderate (10)	Low (35)	Low (10)
	Net Investment Equity	Low (10)	Low (*)	Low (8)	High (2)	Low (12)
	Dependent Student's Income	Moderate (71)	Low (3)	Low (6)	Low (37)	Moderate (6)
	Net Business/Farm Equity	Low (8)	Low (*)	Low (*)	High (1)	Low (13)
	Student's Marital Status	Low (5)	Low (1)	Low (2)	Moderate (10)	Low (15)
	Cash/Savings/Checking	Low (8)	Low (1)	Low (1)	Low (46)	Moderate (7)
	Unusual Medical/Dental Expenses	Low (-2)	Low (*)	Low (*)	Low (23)	Low (14)

in 1 percent.

assets had moderate sensitivity and cash/savings/checking had moderate verifiability. Student's marital status had moderate reliability. Net investment and business/farm equity both were classified as having high reliability. Thus, all seven could be reasonably considered for elimination under this classification.

For the data items in the moderate joint classification, consideration of eliminating them from the Pell formulae becomes a process of dealing with the tradeoffs among measures. Veteran's benefits and elementary and secondary tuition have identical impact across all measures, having low budgetary and distributional impact, moderate sensitivity and verifiability and high reliability. Number in college has moderate budgetary and distributional impact, and sensitivity and low reliability and verifiability. Nontaxable income has moderate budgetary impact, sensitivity and verifiability and low distributional impact and reliability. Net home equity has moderate budgetary and distributional impact, high sensitivity, but low reliability and verifiability.

The remaining items (AGI, social security education benefits, U.S. taxes, family size, and employment expense offset) have the highest impact across the five measures. Within this framework, these items could not be eliminated without a major impact on the program.

The above discussion is an example of how a policymaker might integrate these data given the weighting and classification. Alternative weights could be assigned to each measure, suggesting that some of the measures, such as budgetary impact, are more important than others. In the second example of integrating the scores from the individual measures, we have selected budgetary impact as most important, distributional impact and sensitivity as more important and reliability and verifiability as less important. Thus, we have assigned a weight of three to budgetary impact, a weight of two to distributional impact and sensitivity and a weight of one to reliability and verifiability. Effectively this means that budgetary impact has three times the weight of verifiability, implying greater importance.

Table 9 presents an example of how this differential weighting affects the classification of data elements. One will notice that the classification of the data elements was not affected by differential weighting. The differential weights may, however, affect the decision to eliminate an individual data element within a classification. For example, number in college received moderate classifications on budgetary and distributional impact and sensitivity and low classifications on reliability. Using equal weights, one might choose to eliminate this item. Assigning the differential weights, however, may lead one to reconsider the elimination of the item, since the measures on which the data item received moderate classifications would be assumed to be more important. Greater changes in classification would occur as the difference between the highest and lowest weights increase. This example suggests, however, that classification is relatively unaffected by small changes in weights.

CONCLUSION

This chapter has presented the results of a systematic analysis of the impact of individual data elements designed to provide ED policymakers with the data needed to make informed decisions concerning potential data element reduction options. Each section has presented the results of analyses on an individual measure. The final section presents a framework that policymakers will find useful for integrating these individual analyses, which would be necessary to simultaneously consider the measures. This section also provides two examples of how the framework could be used, employing different weighting schemes. The result is a powerful analytic tool for ED policymakers to develop and evaluate potential data element reduction proposals.

A word of caution should be offered concerning the interpretation of the joint consideration of measures. The analysis assessed the impact of eliminating individual data elements. These results cannot inform policymakers about the cumulative effects of eliminating groups of data elements. The following chapter provides an evaluation of the effects of one such alternative, a five element formula.

TABLE 9

EXAMPLE OF JOINT RANKING OF THE DATA ELEMENTS ASSIGNING DIFFERENTIAL WEIGHTS TO EACH MEASURE

Classification		Budgetary Impact (Weight=3) ($ Million)	Distributional Impact (Weight=2) (%Δ)	Sensitivity (Weight=2) (%Δ in award)	Reliability (Weight=1) (% w/error)	rifiability (Weight=1) (Rank)
HIGH	Adjusted Gross Income	High (1,708)	High (42)	High (1,507)	Low (16)	High (1)
	Social Security Education Benefits	High (276)	Moderate (8)	High (194)	Moderate (5)	Moderate (5)
	Family Size Offset	High (-1,455)	High (51)	High (-65)	Low (22)	Low (16)
	U.S. Taxes Paid	Moderate (-155)	High (15)	High (-85)	Low (14)	High (3)
	Employment Expense Offset	Moderate (-80)	Moderate (10)	Moderate (-27)	Low (18)	High (2)
MODERATE	Net Home Equity	Moderate (117)	Moderate (7)	High (100)	Low (10.7)	Low (11)
	Number in College	Moderate (-100)	Moderate (8)	Moderate (-17)	Low (14)	Low (17)
	Nontaxable Income	Moderate (90)	Low (5)	Moderate (14)	Low (31)	Moderate (8)
	Veteran's Education Benefits	Low (13)	Low (1)	Moderate (12)	High (1)	Moderate (4)
	Elementary and Secondary Tuition	Low (-13)	Low (1)	Moderate (-11)	High (2)	Moderate (9)
LOW	Dependent Student's Net Assets	Low (35)	Low (3)	Moderate (10)	Low (35)	Low (10)
	Net Investment Equity	Low (10)	Low (*)	Low (8)	High (2)	Low (12)
	Dependent Student's Income	Moderate (71)	Low (3)	Low (6)	Low (37)	Moderate (6)
	Net Business/Farm Equity	Low (8)	Low (*)	Low (*)	High (1)	Low (13)
	Student Marital Status	Low (5)	Low (1)	Low (2)	Moderate (10)	Low (15)
	Cash/Savings/Checking	Low (8)	Low (1)	Low (1)	Low (46)	Moderate (7)
	Unusual Medical/Dental Expenses	Low (-2)	Low (*)	Low (*)	Low (23)	Low (14)

*Less than 1 percent.

3

ANALYSIS OF A FIVE DATA ELEMENT FORMULA

The prior chapter presented an analysis of the impact of eliminating individual data from the Pell Grant eligibility and award formulae. This chapter presents an analysis of one proposal to reduce the number of application data elements that are used to calculate Pell awards to five. As described in the Introduction, this analysis is better able to isolate the effects of eliminating data elements by controlling for reporting error. We have controlled for error by conducting analyses of a second simulation using a data base from which error has been eliminated by imputing error patterns found in the Stage III Pell QC data base to the applicant data base. This imputation procedure is presented in detail in Technical Appendix B.

DESCRIPTION OF THE SIMULATIONS

The two simulations conducted by the Division of Policy and Program Development in this analysis—the standard and the error free simulations—are identical with the exception of the data base used. Each simulation consists of three model runs, the first of which develops a baseline measure using the full formula in the 1982-83 program year. Both simulations then eliminate all but five data elements. (Dependency status remains in the formula and is not treated explicitly as a data element.) These are:

- Adjusted gross income
- Federal taxes paid
- Nontaxable income
- Number in household
- Number in postsecondary education.

42

Eliminated from the formula are the following income, asset, and expense data (not necessarily data elements):

- Student/spouse income
- Net home assets
- Net farm and business assets
- Cash, checking, savings
- Net interest assets
- Dependent student's assets
- Offset for unreimbursed elementary and secondary tuition
- Offset for high medical and dental expenses
- Employment expense offset
- Social Security Education Benefits
- Veteran's Education Benefits.

The second run, which uses a five element formula, is used to estimate the adjustments to formula "taxation" rates required to maintain budget neutrality. Budget neutrality was one parameter for analysis specified by ED. Tax rate adjustments are necessary because reducing the formula to five elements causes the budget to increase by approximately $130 million. The tables in Appendix D (Tables D-1 and D-2) display this increase for both data bases when tax rates are not adjusted.

The third run has taxation rates adjusted to maintain budget neutrality (Table 10)[7]. The analysis primarily focuses on the first (full formula) and third (five element with taxation rate adjustments) runs. This analysis explicitly identifies the effects of data element reduction using a standard and "error free" data base while maintaining budget neutrality.

The analysis of both simulations focuses on four policy questions that will assist OSFA policymakers in evaluating data element reduction as a potential corrective

[7] More information concerning the effects of taxation rates can be obtained by consulting **The Pell Grant Formula**, 1982-83, U.S. Department of Education, Office of Student Financial Assistance.

TABLE 10

TAXATION RATES FOR PARENTS' DISCRETIONARY INCOME USING BOTH THE STANDARD AND ERROR FREE SIMULATIONS

Discretionary Income	Standard Taxation Rate	Adjusted Taxation Rate
0 to $ 5,000	11% of discretionary income	13% of discretionary income
$ 5,001 to $10,000	$ 550 + 13% of amount over $5,000	$650 + 15% of amount over $5,000
$10,001 to $15,000	$1,200 + 18% of amount over $10,000	$1400 + 27% of amount over $10,000
$15,001 and above	$2,100 + 25% of amount over $15,000	$2750 + 30% of amount over $15,000

action. The results from these simulations are compared to assess the effects of data element reduction under different simulations. These four questions are:

- How do eligibility and awards change when data elements are reduced to five?
- What are the characteristics of those who gain and lose from the program changes?
- What are the characteristics of newly-eligible recipients?
- What are the characteristics of students who lose eligibility?

These simulations are presented below.

Standard Simulation Using Reported Data

DPPD staff conducted a simulation of the effects of reducing the number of data elements to five using the standard data base (reported data) holding budget constant. The results, organized around the four questions, are as follows:

How do eligibility and awards change?

Generally, analysis of the standard simulation indicates that at the highest level of aggregation, reducing the number of data elements results in very small changes in the number of recipients, distribution of recipients by income strata, and mean award. More specifically, the findings indicate that:

- Although the budget remains approximately constant, the adjustment of taxation rates to maintain a constant budget produces slight increases in the number of recipients by over 50,000 (2 percent) , when the number of data elements is reduced to five (Table 11).
- The proportion of program costs awarded to higher income recipients declines slightly. The mean award decreases to $960 from about $980.
- About 82 percent of those applicants ineligible under the full formula remain so under the reduced formula (Table 12).
- The majority of recipients in most award strata receive the same award (the center diagonal of Table 12).

TABLE 11

COMPARISON OF NUMBER OF RECIPIENTS AND PROGRAM COSTS FOR THE 1982-83 PELL PROGRAM YEAR UNDER THE FULL AND FIVE DATA ELEMENT FORMULAE USING STANDARD REPORTED DATA

	Full Formula				Five Element Formula			
	ESTIMATED		COMPUTED EXPECTED AWARD		ESTIMATED		COMPUTED EXPECTED AWARD	
	NUMBER OF RECIPIENTS	% OF TOTAL RECIPIENTS	TOTAL COST	PERCENT OF TOTAL PROGRAM COST	NUMBER OF RECIPIENTS	% OF TOTAL RECIPIENTS	TOTAL COST	PERCENT OF TOTAL PROGRAM COST
TAL	2527705	100	2480968116	100	2579411	100	2488530288	100
TAL INCOME								
OR LESS	147501	6	161803469	7	151202	6	165421402	7
4,000	623404	25	720283983	29	629747	24	717598467	29
001-7,500	527031	21	564716739	23	536301	21	571251347	23
501-10,000	254096	10	277498975	11	246640	10	285357758	11
,001-12,000	179806	7	184351905	7	191052	7	189215613	8
,001-15,000	233699	9	215216344	9	243314	9	221557530	9
,001-20,000	267650	11	212883444	9	305103	12	217353758	9
,001-25,000	162665	6	97387244	4	143389	6	85392321	3
,001-30,000	81508	3	37595849	2	63622	2	26187496	1
,000 OR MORE	45342	2	17227164	1	25041	1	9202535	0

TABLE 12

COMPARISON OF PELL AWARDS FOR THE 1982-83 PROGRAM YEAR UNDER FULL AND FIVE DATA ELEMENT FORMULAE USING STANDARD REPORTED DATA

Full Formula

	TOTALS		0	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1,000	1,001-1,100	1,101-1,200	1,201-1,300	1,301-1,400	1,401-1,500	1,501-1,600	1,601-1,700
	TOTALS		TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS
NUMBER	3601522		1078810	9400	107020	142762	103007	167175	103340	140047	149584	103440	202150	155522	196170	180020	126231	143434	89800	322752
%	100.00		29.95	0.26	2.97	3.96	2.86	4.64	2.87	3.89	4.15	5.09	5.61	4.32	5.45	5.02	3.50	3.90	2.49	0.96
	TOTALS		COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN	COLUMN
	NUMBER	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
Five Element Formula																				
0	1026111	28.4	81.60	84.20	47.54	33.30	16.57	11.49	3.13	1.50	0.25	0.21	0.13	.	0.01	0.00	0.01	.	.	0.02
1-100	11721	0.3	0.13	26.00	3.25	3.07	1.12	0.12	0.07	0.03	0.00	.	0.01	.	.	.	.	.	.	.
101-200	104779	2.9	2.15	5.05	23.94	9.99	9.03	12.28	4.09	3.09	0.97	0.89	0.17	0.01	0.03	0.02	0.01	0.01	.	.
201-300	147066	4.0	2.03	3.00	4.61	31.66	9.51	10.84	7.99	15.92	1.94	1.33	0.91	0.02	0.10	0.05	0.00	.	0.02	0.00
301-400	108550	3.01	1.35	2.33	2.60	2.36	47.66	6.32	4.97	6.70	5.40	0.02	0.94	0.07	0.06	0.05	0.02	.	.	.
401-500	165520	4.60	2.00	2.41	3.46	2.01	3.51	43.12	6.36	9.77	6.60	0.55	4.59	0.59	1.14	0.39	0.04	0.02	0.07	.
501-600	101230	2.81	0.72	2.33	0.84	1.40	1.21	1.29	65.37	3.77	2.20	1.64	2.90	0.14	0.35	0.34	0.02	0.02	.	0.00
601-700	151100	4.20	1.57	1.54	3.34	2.76	1.96	2.07	2.00	44.54	5.73	5.54	10.45	2.65	4.19	1.07	0.00	0.04	0.03	.
701-800	153700	4.27	0.91	0.67	1.34	1.50	1.29	1.90	1.43	3.13	40.85	3.53	3.25	1.77	3.92	1.00	0.74	0.15	0.03	0.01
801-900	107979	5.22	0.97	.	1.69	1.00	1.47	1.60	0.93	2.11	2.37	72.40	4.54	2.30	2.32	4.35	0.45	0.36	0.04	0.00
901-1,000	212252	5.89	1.37	.	2.07	3.67	1.63	2.21	1.66	3.50	1.49	1.02	45.57	4.39	5.23	7.63	5.97	1.20	1.02	0.03
1,001-1,100	154030	4.20	0.39	.	0.55	0.64	0.66	0.56	0.62	1.00	0.71	0.07	0.07	84.24	1.27	1.05	1.23	4.04	0.13	0.01
1,101-1,200	203427	5.65	0.90	.	1.32	1.30	1.07	1.23	0.97	1.77	0.81	0.96	2.19	0.00	74.79	4.00	0.51	2.53	0.60	0.17
1,201-1,300	187935	5.19	0.84	.	0.90	1.17	0.04	1.45	0.54	1.61	0.90	0.02	1.44	0.30	1.31	74.92	2.04	7.70	2.53	0.40
1,301-1,400	129411	3.59	0.66	.	1.03	0.67	0.41	0.01	0.39	0.89	0.51	0.48	0.70	0.33	0.95	1.45	49.03	2.00	10.41	0.60
1,401-1,500	148377	4.12	0.60	.	0.60	0.70	0.40	0.75	0.12	0.76	0.63	0.40	0.61	0.13	0.52	0.07	2.32	77.00	2.63	2.79
1,501-1,600	88574	2.46	0.30	.	0.20	0.36	0.29	0.32	0.15	0.41	0.20	0.23	0.37	0.11	0.34	0.40	1.04	1.20	47.91	4.60
1,601-1,700	320573	0.96	0.54	.	0.49	0.57	0.49	0.47	0.13	0.42	0.33	0.10	0.40	0.04	0.44	0.37	1.95	2.35	7.69	91.27
TOTALS	3601522	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Note On Interpreting This Table: The above table indicates the percentage of recipients who receive awards under the reduced formula that are the same, greater or less than those received under the full formula. The center diagonal lines from top left to bottom right highlight the percentage of recipients within each award range (e.g., $501 - 600) whose award was unchanged under the reduced formula. For example, about 65% received an award of between $501 and $600 under both formulae. About 26% received less and about 9% received more under the reduced formula. Two percent of those who received an award between $501 and $600 under the full formula received between $601 and $700 under the reduced formula. The areas in the upper right and lower left set off by single diagonal lines indicate the greatest changes in awards.

Technical Note: The totals in this simulation do not equal the actual number of applicants because a participation rate (or no show rate) has been applied to all applicants by adjusting the sampling weight of each applicant. The result is a reduction in the overall number of applicants to more accurately reflect the number that become recipients. The number of recipients is accurate.

- Of those receiving the maximum awards--the neediest students--91 percent continue to receive the maximum award and 99 percent receive awards within $200 of the maximum. Of those recipients receiving an award of more than $1400, 92 percent receive more than $1400 under the f[illegible]e element formula.

- Of those receiving the lowest awards (not greater than $400) 49 percent continue to receive an award not greater than $400.

What Are the Characteristics of Those Who Gain and Lose?

In general, the following patterns describe those applicants who have their awards increased (gainers) or decreased (losers):

- Most gainers are clustered in the middle of the award range; students receiving smaller awards (below $500) are more likely to lose under the reduced formula than those receiving the higher awards. The neediest students, those receiving the highest awards, are least likely to lose significant amounts. Relatively few applicants gain or lose extremely large amounts (upper right and lower left sections of Table 12).

- Of those whose awards increase, 66 percent increase by less than $600, 25 percent increase by $600 - $1,200 and 9 percent by more than $1,200.

- Those gaining less than $600 had a mean AGI of $12,700 and mean net assets of almost $40,000; those gaining $600 - $1,200 had a mean AGI of $12,500 and mean net assets of $54,000; and those gaining over $1,200 had a mean AGI of $9,000 and mean net assets of $92,000.

- Of those recipients whose awards decrease, almost 98 percent decrease by less than $600; about 2 percent decrease by $600 - $1,200 and less than .1 percent by more than $1,200.

The following data summarize the percentage of Pell Grant recipients who gain, lose, and stay the same (within $50) by specific demographic and financial characteristics under the five data element formula when compared with the current formula.

Characteristics	Percentage Who Receive a Smaller Award	Percentage Who Receive the Same Award (± $50)	Percentage Who Receive a Larger Award
• All Applicants	16	73	11
• Dependent Students with Family Size 4 and Under	20	63	17
• Dependent Students with Family Size 5 and Over	19	66	15

Characteristics	Percentage Who Receive a Smaller Award	Percentage Who Receive the Same Award (± $50)	Percentage Who Receive a Larger Award
• Independent Students	11	86	3
• Families with 1 in Postsecondary Education	14	77	9
• Families with more than 1 in Postsecondary Education	21	63	16
• Dependent Students with Net Home Value under $10,000	26	65	9
• Dependent Students with Net Home Value over $10,000	16	64	20
• Dependent Students with Family Investments under $10,000	21	64	15
• Dependent Students with Family Investments Over $10,000	8	69	23
• Dependent Students with Family Business/Farm Value Under $10,000	20	65	15
• Dependent Students with Family Business/Farm Value Over $10,000	12	59	29
• Families with No Nontaxable Income	18	71	11
• Families with Some Nontaxable Income	13	76	11
• Dependent Students with No Extraordinary Family Medical/Dental Expenses	17	71	12
• Dependent Students with Any Extraordinary Medical/Dental Expenses	21	61	18
• Student Enrolled Full-Time	16	72	12
• Student Enrolled Less Than Full-Time	13	82	5

From these data we can conclude that:

- Almost three-quarters of all applicants would receive the same award under the reduced formula as under the full formula; one-quarter would receive a different award.
- The vast majority (86 percent) of independent students are unaffected by data element reduction.
- Students who fare better than average under data element reduction as expected are those from families with higher home equity, larger investments, businesses, or farms. These wealth elements are not considered in the reduced data formula.
- Students enrolled less than full-time, reflecting a high proportion of independent students, are less likely to be affected by data element reduction than are full-time students.

What Are the Characteristics of Newly Eligible Recipients?

- An estimated 200,000 applicants who are ineligible under the full formula would become eligible under the reduced formula.
- Of these newly eligible recipients, half would receive an award of less than $600, one-third would receive between $600 and $1,200, and one-sixth over $1,200.
- Those newly eligible recipients gaining less than $600 had a mean AGI of $20,000 and mean net assets of $57,000; those gaining awards of between $600 and $1,200 had a mean AGI of $15,000 and mean net assets of $61,000; and those gaining awards in excess of $1,200 had a mean AGI of $9,000 and mean net assets of $97,000.

What Are The Characteristics Of Students Who Lose Eligibility?

- Slightly less than 150,000 students who received awards under the full formula become ineligible under the reduced formula.
- Of the 360,000 who received an award of less than $401 under the full formula, 33 percent became ineligible. Almost no one among the 1.2 million students who received in excess of $1,000 under the full formula became ineligible under the reduced formula.
- Those students who lost an award of less than $600 had a mean AGI of $24,000 and mean net assets of $14,000; those who lost an award between $600 and $1,200 had a mean AGI of $22,000 and mean net assets of $7,000 and those who lost an award in excess of $1,200 had a mean AGI of $12,000 and mean net assets of $8,000.

"Error Free" Simulation

The Division of Policy and Program Development conducted a second simulation of the impact of a reduced formula using a data base to which "best values" were imputed. This imputation effectively removed reporting error from the data base and permitted a more accurate measurement of the effects of data element reduction as distinct from the elimination of error. This simulation focuses on the same four questions as the standard simulation.

How Do Eligibility and Awards Change?

This simulation also indicates that at the highest level of aggregation, reducing the number of data elements, using an error free data base, results in even smaller changes in the number of recipients, distribution of recipients by income strata, and no change in mean award. More specifically, the findings indicate:

- Maintaining approximate budget level results in a negligible increase in recipients, about 11,000 or less than .5 percent (Table 13).
- The proportion of program costs awarded to low income recipients increases slightly and the proportion awarded to high income recipients decreases.
- The mean award of $940 is unchanged.
- Over 86 percent of those 1.2 million ineligible applicants under the full formula remain ineligible under the reduced formula (Table 14).
- The majority of recipients in most award strata receive the same award (the center diagonal of Table 14).
- Of the 250,000 students receiving maximum awards--the neediest students--90 percent continue to receive the maximum and 98 percent receive within $200 of the maximum. Of the 480,000 students receiving more than $1,400, 92 percent continue to receive in excess of $1,400.
- Just under 50 percent of the 350,000 students who received $400 or less under the full formula continue to receive an award of $400 or less. Thirty-six percent of the students who originally received $400 or less become ineligible.

What Are the Characteristics Of Those Who Gain and Lose?

In general, the following patterns describe those applicants whose awards

TABLE 13

COMPARISON OF NUMBER OF RECIPIENTS AND PROGRAM COSTS FOR THE 1982-83 PELL PROGRAM YEAR UNDER THE FULL AND FIVE DATA ELEMENT FORMULAE USING ERROR FREE DATA

	Full Formula				Five Element Formula			
	ESTIMATED		COMPUTED EXPECTED AWARD		ESTIMATED		COMPUTED EXPECTED AWARD	
	NUMBER OF RECIPIENTS	% OF TOTAL RECIPIENTS	TOTAL COST	PERCENT OF TOTAL PROGRAM COST	NUMBER OF RECIPIENTS	% OF TOTAL RECIPIENTS	TOTAL COST	PERCENT OF TOTAL PROGRAM COST
TOTAL	2421948	100	2273058334	100	2433397	100	2285124542	100
TOTAL INCOME								
0 OR LESS	67331	3	72103607	3	68046	3	73489603	3
1-4,000	511594	21	568188474	25	515161	21	574838031	25
4,001-7,500	557096	23	574611825	25	568451	23	592479228	26
7,501-10,000	267539	11	277509303	12	277666	11	[illegible]4123875	13
10,001-12,000	202377	8	200203886	9	211916	9	2059[illegible]467	9
12,001-15,000	247[illegible]6	10	2160[illegible]173	10	255349	10	223922368	10
15,001-20,000	309228	13	234960120	10	324934	13	221038159	10
20,001-25,000	164676	7	90566541	4	14958[illegible]	6	78758883	3
25,001-30,000	67109	3	28963111	1	46103	2	19614494	1
30,000 OR MORE	27973	1	9912294	0	16230	1	6266436	0

TABLE 14

COMPARISON OF PELL AWARDS FOR THE 1982-83 PROGRAM YEAR UNDER FULL AND FIVE DATA ELEMENT FORMULAE USING ERROR FREE DATA

Full Formula

	TOTALS		0	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1,000	1,001-1,100	1,101-1,200	1,201-1,300	1,301-1,400	1,401-1,500	1,501-1,600	1,601-1,700
	TOTALS		TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS	TOTALS
NUMBER	3612752		1190004	10700	104130	136001	106030	173036	94493	140600	164265	186516	228002	172920	164140	145309	111490	128051	100734	251747
%	100.00		32.96	0.30	2.00	3.76	2.96	4.01	2.63	3.89	4.55	5.16	6.31	4.79	4.66	4.02	3.09	3.54	2.79	6.97
	TOTALS NUMBER	%	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %	COLUMN %
Five Element Formula																				
0	1179355	32.64	86.37	53.63	49.22	36.70	10.63	10.51	2.05	1.92	0.23	0.20	0.05	0.02	0.04	0.02	0.04	.	.	0.01
1 100	11931	0.33	0.07	23.05	3.41	2.26	1.44	0.12	0.02	0.02	0.02	.	0.06	.	.	.	.	.	.	.
101-200	96020	2.60	1.47	4.76	22.93	10.05	9.40	12.26	2.90	3.49	0.00	0.15	0.11	0.03	0.03	6.05	6.03	0.03	6.01	.
201-300	139662	3.04	1.95	3.30	6.34	31.06	11.05	12.33	9.36	12.95	1.90	1.52	0.65	0.11	0.69	0.63	6.00	0.02	0.03	0.61
301-400	110753	3.67	1.05	4.22	2.52	2.76	47.13	7.11	5.03	10.60	2.75	0.73	0.95	0.04	0.63	6.05	.	6.61	0.01	.
401-500	171246	4.74	1.71	3.79	2.60	3.30	3.36	44.02	6.91	12.22	9.67	7.41	2.59	0.59	1.22	0.44	0.04	0.66	.	.
501 600	94197	2.61	0.50	5.22	0.90	1.00	1.33	1.42	44.56	3.70	2.50	2.10	2.39	0.27	0.26	0.36	0.06	.	.	0.00
601-700	146014	4.04	1.34	2.49	2.61	2.60	1.72	2.64	2.31	42.33	6.37	6.13	9.09	2.70	3.25	1.67	0.91	0.11	0.02	0.63
701 800	164369	4.55	0.02	.	1.19	1.40	0.93	1.40	1.20	2.16	40.36	3.71	4.00	1.00	4.46	1.19	6.79	0.16	6.62	.
801-900	191952	5.31	0.74	0.29	1.24	1.75	1.42	1.40	1.12	2.42	3.12	73.46	5.20	2.16	4.12	2.93	6.71	0.90	0.06	6.04
901 1,000	228530	6.33	1.07	0.27	1.93	2.09	1.11	1.99	1.40	2.54	1.32	1.03	48.61	4.62	6.34	9.40	4.22	0.91	6.04	0.01
1,001-1,100	160104	4.66	0.33	.	1.12	0.77	0.21	0.05	0.65	1.09	0.42	0.66	1.23	85.62	1.54	1.13	6.95	0.10	6.65	0.02
1,101-1,200	169526	4.69	0.55	.	2.16	1.67	0.60	1.07	0.53	1.25	0.96	0.65	1.35	1.37	74.72	6.05	9.21	1.51	6.74	0.15
1,201-1,300	154077	4.26	0.62	.	1.26	0.74	0.55	0.91	6.50	1.35	0.50	0.47	1.01	6.30	2.69	73.26	9.90	0.54	1.79	0.62
1,301 1,400	112991	3.13	6.49	.	1.40	0.60	0.22	0.66	6.41	0.69	0.36	0.36	0.70	0.15	0.05	1.64	60.13	7.50	7.74	0.74
1,401-1,500	123046	3.43	0.36	.	6.53	0.46	0.20	0.61	0.13	0.40	0.30	0.36	0.70	0.16	0.41	1.10	2.50	71.90	9.00	2.60
1,501 1,600	83059	2.30	6.21	.	0.22	0.30	0.17	0.26	0.03	0.36	0.13	0.09	0.30	0.03	0.31	0.37	0.04	3.24	56.00	5.65
1,601-1,700	266307	7.37	0.35	.	0.32	0.19	6.29	0.34	6.67	0.34	0.23	6.09	0.18	0.03	0.23	0.64	1.50	5.00	22.04	90.05
TOTALS	3612752	100.00	160.06	160.06	160.00	100.00	160.00	100.00	100.00	100.00	100.00	1?0.00	166.00	160.00	100.00	100.90	106.60	166.00	160.06	106.00	100.00

Note On Interpreting This Table: The above table indicates the percentage of recipients who receive awards under the reduced formula that are the same, greater or less than those received under the full formula. The center diagonal lines from top left to bottom right highlight the percentage of recipients within each award range (e.g., $501 - 600) whose award was unchanged under the reduced formula. For example, abo : 65% received an award of between $501 and $600 under both formulae. About 27% received less and about 8% received more under the reduced formula. Two percent of those who received an award between $501 and $600 under the full formula received between $601 and $700 under the reduced formula. The areas in the upper right and lower left set off by single diagonal lines indicate the greatest changes in awards.

Technical Note: The totals in this simulation do not equal the actual number of applicants because a participation rate (or no show rate) has been applied to all applicants by adjusting the sampling weight of each applicant. The result is a reduction in the overall number of applicants to more accurately reflect the number that become recipients. The number of recipients is accurate.

increased or decreased.

- Table 14 indicates that most students whose awards increased are clustered in the middle of the award range; those receiving a smaller award ($500) under the full formula are most likely to receive a smaller award under the reduced formula. The neediest students, those receiving the highest awards, are least likely to have their awards decrease signficiantly. Relatively few applicants gain or lose extremely large amounts (the upper right and lower left of Table 14).

- Of those students whose awards increased, 72 percent increased by less than $600, 22 percent increased by $600 - $1,200 and 6 percent by more than $1,200.

- Those gaining less than $600 had a mean AGI of $13,000 and mean net assets of $34,000; those gaining between $600 and $1,200 had a mean AGI of $14,000 and mean net assets of $57,000 and those gaining over $1,200 had a mean AGI of $10,000 and mean net assets of $90,000.

- Of those whose awards decreased, slightly less than 98 percent decreased less than $600, about 2 percent decreased between $600 and $1,200 and less than .1 percent decreased more than $1,200.

- Those students losing less than $600 had a mean AGI of almost $17,000 and mean net assets of $12,000: those losing between $600 and $1,200 had a mean AGI of $14,000 and mean net assets of $6,000; those losing more than $1,200 had a mean AGI of about $12,000 and mean net assets of $3,000.

The following data summarize the percentage of Pell Grant recipients who gain, lose, and stay the same (within $50) by specific demographic and financial characteristics under the five data element formula when compared with the current formula using error free data in both runs.

Characteristics	Percentage Who Receive a Smaller Award	Percentage Who Receive the Same Award (± $50)	Percentage Who Receive a Larger Award
• All Applicants	18	72	10
• Dependent Students with Family Size 4 and Under	22	64	15
• Dependent Students with Family Size 5 and Over	20	66	14
• Independent Students	13	85	3
• Families with 1 in Post-secondary Education	17	74	9

Characteristics	Percentage Who Receive a Smaller Award	Percentage Who Receive the Same Award (± $50)	Percentage Who Receive a Larger Award
• Families with more than 1 in Postsecondary Education	22	63	14
• Dependent Student with Net Home Value under $10,000	26	64	10
• Dependent Student with Net Home Value over $10,000	18	65	17
• Dependent Students with Family Investments under $10,000	22	64	14
• Dependent Students with Family Investments Over $10,000	12	69	19
• Dependent Students with Family Business/Farm Value Under $10,000	21	65	14
• Dependent Students with Family Business/Farm Value Over $10,000	16	60	24
• Families with No Nontaxable Income	19	72	9
• Families with Some Nontaxable Income	18	71	11
• Dependent Students with No Extraordinary Family Medical/Dental Expenses	18	70	12
• Dependent Students with Any Extraordinary Medical/Dental Expenses	22	62	15
• Student Enrolled Full-Time	18	71	11
• Student Enrolled Less Than Full-Time	15	79	6

From this table we can conclude that:

- Almost three-quarters of all applicants would receive the same award under the reduced formula as under the full formula.
- The vast majority (85 percent) of independent students are unaffected by data element reduction.
- Students who fare better than average under data element reduction as expected are those from families with higher home equity, larger investments, businesses, or farms. These wealth elements are not considered in the reduced data element formula.
- Students enrolled less than full-time, reflecting a high proportion of independent students, are less likely to be affected by data element reduction than are full-time students.

What Are the Characteristics Of Newly Eligible Recipients?

- An estimated 162,000 applicants who were ineligible under the full formula would become eligible under the reduced formula.
- Approximately 46 percent of these newly eligible recipients would receive $600 or less, 36 percent between $601 and $1,200 and 18 percent more than $1,200.
- Newly eligible receipients who would receive an award of less than $600 had a mean AGI of $19,000 and mean net assets of $58,000; those who would receive between $600 and $1,200 had a mean AGI of 15,000 and mean net assets of $67,000; those who would receive over $1,200 had a mean AGI of $10,000 and mean net assets of $95,000.

What Are The Characteristics of Students Who Lose Eligibility?

- An estimated 151,000 students who were eligible under the full formula would lose eligibility under the reduced formula.
- Of those 151,000 who lose eligibility, 92 percent lose awards of less than $600, slightly less than 8 percent lose awards between $600 and $1,200 and less than 1 percent lose awards of over $1,200.
- Virtually all of the neediest students, those receiving maximum awards, remain eligible.
- Those students who lost less than $600 had a mean AGI of $22,000 and mean net assets of $16,000. Those losing between $600 and $1,200 had a mean AGI of $19,000 and mean net assets of $9,000. Those losing in excess of $1,200 had a mean AGI of $17,000 and mean net assets of $6,000.

FINDINGS

The simulations presented in the prior sections of this chapter result in several outcomes. The first of these is a more thorough understanding of the budgetary and distributional effects of reducing the number of data elements that are used to calculate Pell eligibility and awards to five.

The second outcome is the development of a thorough description of the comparative effects of data element reduction controlling for error. These comparative budgetary and distributional effects can be expressed on several levels. The data indicate the following general findings:

- The greatest differences in the impact of data element reduction using the two data bases are evident at the aggregate level including program costs and number of recipients. Results are fairly similar across many dimensions on a more detailed level.
- Use of an error free data base in simulating the effects of data element reduction dampens the increase in recipients and slightly increases the budgetary impact.

More specifically, a comparison of the two simulations indicates the following:

- The error free five element formula with tax rate adjustments results in a level of recipients that is 142,000 students less than the standard simulation.
- The baseline budget for the error free simulation is $215 million dollars less than the baseline budget for the standard simulation ($2.48 billion).
- The net increase in program costs for an error free reduced formula without tax rate adjustments ($149 million) is slightly larger than for the standard simulation of a reduced formula without tax rate adjustments ($130 million, see Appendix D, Tables D-1 and D-2).
- Program costs for an error free simulation of a five element formula without tax rate adjustment are equal to the baseline program costs of about $2.48 billion, suggesting that when error is eliminated, no increase in taxation rates is necessary. (See Appendix D, Table D-2.)
- The average award in the error free simulations is unchanged under the reduced formula, while the average award drops slightly in the standard simulation.

On a more detailed level the simulations produce different results on the following dimensions:

- More students receiving low awards ($500 or less) continue to receive such awards under the standard simulations.

- More students receive lower increases ($600 or less) under the error free simulation.

Differences on other dimensions between the simulations are minimal (e.g., within 2 to 3 percentage points) and mixed.

CONCLUSION

This chapter has presented the results of two simulations of reducing the number of data elements in the Pell need analysis formula to five. These simulations have advanced general understanding of the effects of data element reduction on an aggregate and an individual level.

The second of these simulations was conducted with a data base from which error has been eliminated. This simulation permitted modeling the joint effects of eliminating error as well as reducing the number of data elements in the Pell formulae for the first time. A comparison of these simulations has permitted a better understanding of the implication of error on the prevalent assumptions concerning data element reduction and the differences relating to specific effects.

A word of caution should be offered concerning the interpretation of the findings. These findings are subject to the same caveats concerning the static nature of the data base discussed in the Introduction. Perhaps a more important caveat however, relates to the analyses. We have designed these analyses as an evaluation not as a forecast of likely policy outcomes. An example of this difference is evident in the assumptions underlying the imputation of error to the data base and the error free simulation. We assume in this imputation and simulation that all error found in Pell QC Stage III is eliminated—even from the remaining data elements. Clearly, this is an unlikely assumption for a policy forecast. However, it is fundamental to our analysis from a research perspective and has produced valuable results.

TECHNICAL APPENDIX A

DESCRIPTION OF THE MODEL AND DATA BASE

The program-wide simulations of full and five element formulae conducted for this report have been produced from the official ED simulation model (the applicant-based model) with which the Pell Grant Branch, DPPD, produced the data tapes for analysis.

The applicant-based model is a micro-model of the Pell Grant Program designed to simulate for ED policymakers changes in awards and recipients under different Pell program parameters. The model uses a weighted sample of approximately 160,000 actual Pell applicants. This data base was used both in the program-wide simulation and the assessment of individual data elements.

The model computes a Student Aid Index, or eligibility index for each applicant using the Pell Grant family contribution schedule. It applies an imputed cost of attendance and enrollment status for each applicant and computes an expected award. Finally, the model applies a "show up rate" or estimation of the number of eligible applicants who will submit Student Aid Reports to postsecondary institutions and receive Pell Grants. The sample of applicants is weighted to produce estimates for the population of applicants and recipients.

The Pell Grant Branch, DPPD, has produced several program-wide simulations of the 1982-83 academic year for this analysis. The baseline simulations, which replicate the 1982-83 year, have the following characteristics:

- The 1981-82 data base aged to represent 1982-83 applicant data
- 1982-83 Pell Grant Program parameters
    - $1,800 legislative maximum award/$1,800 maximum award
    - "Taxation rates" on discretionary income of 11, 13, 18, and 25 percent for dependent students increasing by income levels; 25 percent for married independent applicants and 33 percent for single independent applicants with a family size of one
    - Resource protection of $25,000 for home and an additional $25,000 for other investments

- All awards were reduced by about 6 percent to reflect validation savings. Therefore, the effective maximum award is less than $1700 and the minimum award is less than $100.

- A participation or no show rate stratified by income, was applied to all applicants to estimate the number of eligible recipients who actually receive Pell Grants. This accurately estimates the number of recipients, but reduces the overall number of applicants below actual levels.

TECHNICAL APPENDIX B

IMPUTATION OF STAGE III ERROR PATTERNS TO THE ED APPLICANT DATA BASE

This appendix describes the statistical techniques used to assign "best values"* to the applicant file. The purpose of the assignment procedure was to make possible a statistical simulation of the effects of program error rates on alternative eligibility formula. Statistical procedures used to assign best values were designed to reproduce the patterns of reporting errors discovered in Stage III of the Pell Grant Quality Control Project. This appendix consists of two parts: general approach and imputation procedures.

GENERAL APPROACH

The selection of a procedure with which to most accurately impute best values to the ED applicant data base received much attention and thought, and several approaches were considered and rejected before finally selecting a suitable approach. The objective of the selection process was to maximize the accuracy of the imputation. In order to do so it would be necessary to capture those characteristics that were the greatest predictors of the probability and level of error for any single data element reported value (zero/non-zero), dependency status, income, and error on certain other variables. The approaches considered included:

- Statistical matching
- Regression
- Simultaneous interactions
- "Cold decking"/ratio estimation
- "Cold decking"/regression

One of the most promising and yet straightforward approaches considered was statistical matching. Statistical matching is similar in approach to the commonly used

---

*"Best values," as used in this context, refers to application data values that have been determined to be correct through a variety of data collection techniques used in the Pell Grant Quality Control project.

procedure of exact matching, which matches records from one source with records from another source using identifiers, such as social security number, that* enable the linkage of data from two discrete sources. Statistical matching links records from one data source with a second, similar source by minimizing a specified distance function. (Radner, et al., 1980) Statistical matching is widely used in the preparation, manipulation, and analysis of large scale data bases, for example Census surveys. (Radner, 1983) Matching is often used to impute or assign missing data values to cases on one data base (a recipient) by searching a second data base (a donor) and identifying a donor case that is closest to the case (a recipient) across specified dimensions (e.g., other data values or characteristics) and assigns the value of the item from the donor to the recipient case.

Two types of matching are commonly recognized. The first is unconstrained matching, which places no restrictions on the number of records that are matched from the recipient to the donor file. (Okner, 1972) This approach has several weaknesses, which resulted in our rejecting it as an acceptable approach.

With unconstrained matching both the mean and standard deviation of the estimated variables in the recipient file may differ from the corresponding statistics in the donor file. Unconstrained statistical matching has the advantage of permitting the closest possible match for each recipient record, but at the cost of increasing the sample variance of estimators involving the estimated variables. An unconstrained match amounts to taking a simple random sample with replacement of the records in the donor file. Thus, the distributions of the imputed variables added to the recipient file are distributions of the selected sample rather than the distributions as observed in the recipient file. (Rogers, 1984) For these reasons, we rejected unconstrained matching as an approach to error imputation.

The second type of statistical matching, constrained matching, held more promise as a method. (Barr and Turner, 1980) Constrained matching ensures that each donor file record is matched with a recipient file record by duplication of recipient file records, if necessary. The advantages of a constrained match are that the multivariate distribution of the imputed variables identically match the distribution in

---

*The reader is cautioned not to confuse the concept of donor and recipient used here with the Pell Grant Recipient file and the Pell Grant Applicant file.

the donor file as do the mean and standard deviation. A disadvantage includes the limitation that matched pairs (from both files) potentially differ more with respect to common values than an unconstrained match. The most significant disadvantages are that procedures that minimize the differences between paired cases require considerable computer time, particularly for large data sets, and potentially result in an expanded data set. (Rogers, 1984) This posed serious time, resource, and computational problems, and led to the rejection of this approach.

Another approach to imputation considered was regression. Regression would allow extrapolation of error data beyond the recipient file, a key issue since the applicant file contains data values in excess of the recipient file (e.g., AGI). This, however, was rejected because it would assign a small amount of error to all cases and would not capture the incidence of error and the full impact of this error on individual eligibility and awards.

A procedure of mapping the simultaneous interactions of all errors was considered. This would precisely replicate the error patterns including the level and interaction among errors. It was not considered feasible, since the complexity would have outstripped the computer resources and quickly exhausted the degrees of freedom on the Stage III file. Allowing interaction among the 18 variables, zero and non-zero reported value, error and no error, yields over 68 billion (4^{18}) combinations.

Thus, we considered and adopted a "cold decking" process for cases without dependency status error that stratified the Stage III file on reported value (zero/non-zero), dependency status and income. The probability of error was computed for each stratum. The issue of estimating best values was more difficult. We considered a ratio estimator that, not unlike a regression coefficient, would permit extrapolation of best values beyond the range of recipient reported values. The ratio estimator had two flaws. First, and perhaps most serious, a ratio estimator is inappropriate and ineffective with zero reported values (since zero multiplied by anything is zero), and error patterns were highly dependent on reported value (zero/non-zero).

The ratio estimator also limited the prediction of best value of a single variable to the reported value of that variable and could not account for simultaniety of errors. Because of these limitations we replaced the ratio estimator with multivariate regression models, although we continued to use a "cold decking" procedure stratified in income, reported values, and dependency status. This multivariate regression allowed us to control for the simultaniety of related errors as well as zero/non zero reported values. This is described below.

The cold decking technique employed to assign an application to an error status is currently used by Vital Statistics for estimating out-of-wedlock birth rates, by NCES in its primary and secondary school surveys, and by NCHS for its fetal surveys. Formal statistical analyses of the cold-deck approach can be found in Schaible (1979), Brewer (1979) and Oh and Scheuren (1981).

Under the cold-deck approach the applicant file was first stratified into eight groups:

- Dependent students with total family incomes up to $8,000
- Dependent students with total family incomes between $8,000 and $15,000
- Dependent students with total family incomes between $15,000 and $20,000
- Dependent students with total family incomes over $20,000
- Independent students with incomes up to $2,000
- Independent students with incomes between $2,000 and $4,000
- Independent students with incomes between $4,000 and $8,000
- Independent students with incomes over $8,000

Probabilities for various combinations of error patterns for each strata were estimated from Stage III verified student data. A pattern was defined by the presence or absence of error on each of 18 verified application items.

The patterns were found to depend on whether the reported value was zero. Each variable was subset into zero and non-zero subgroups. For each variable within a stratum there are then four possible events:

- Reported value zero, no error
- Reported value zero, error
- Reported value not zero, no error

71

- Reported value not zero, error

As previously discussed, allowing interaction among the 18 variables, which would exactly reproduce the Stage III error patterns including simultaniety, yields over 68 billion (4^{18}) possible error patterns for each stratum. To reduce the complexity of the error patterns, several assumptions were made based on simultaneous error patterns found in Stage III data. The presence of error on adjusted gross income, nontaxable income, and net home value were assumed to be dependent of each other, but independent of the presence of error on all other data items. Similar relationships were assumed for family size and number in college and for dependent student's income and dependent student's assets. The presence of error on the remaining 11 data items was assumed to be independent of the presence of error on all other data items. Thus, the number of error patterns within each stratum was reduced to 140 ($4^3 + (2 \times 4^2) + (11 \times 4)$).

Error patterns were assigned to applications with probabilities proportional to their occurrence within the strata. For every variable in the pattern assigned that contained no error, the reported value was assumed to be the best value. For variables assigned to an error status the best value was computed as a linear function of the reported value and other variables shown in Stage III to be predictive error values. The formula used was:

$$T = A_i + B + E_i$$

where:

T is a nx1 vector imputed best (true) values

B is a px1 vector of coefficients associated with application variables and an intercept term and estimated using OLS procedures with Stage III data

A_i is a nxp matrix of application values predictive of true values and including the reported value on the variable being imputed

E_i is a nx1 vector of random, normal deviates with an expectation of 0 and a variance equal to the observed residual variance from the Stage III data.

A separate equation was estimated for each of the 18 variables to be imputed in each of the 8 strata for a possible total of 144 equations. Strata were collapsed for some variables due to small degrees of freedom.

Given assumptions of linearity within the parameters, a normal distribution of errors and E(B/Recipients) = E(B/applicants) then Ericson (1969), Royall (1970) and Cochran (1977) have shown that $A_i\bar{B}$ is the maximum likelihood estimate of T within a stratum. We added E to A_iB to reproduce the observed within strata variance while preserving the unbiased expectation of T_i;

Because $E(E_i) = O$ and given the assumptions above;

$E(A_iB) = T_i$

Therefore $E(A_iB + E_i) = T_i$

Regression models for family size and number in college did not provide sufficient predictive results. The joint distribution of best family size and best number in college conditioned on reported dependency status, reported family size, and reported number in college was determined for the recipient data base. This distribution was then imputed to the applicant data base. The following example illustrates this procedure for a selected combination of dependency status, reported family size, and reported number in college.

Dependent Students Reporting Family Size of Four and Two Enrolled in Postsecondary Education

Distribution of Best Values

	Number in College			
Family Size	1	2	3	Total
2	2.4	1.2	0	3.6
3	13.7	4.7	1.2	19.6
4	8.3	59.6	.6	68.5
5	.6	5.9	0	6.5
6	0	.6	.6	1.2
7	0	0	.6	.6
Total	25	72	3	100

Whenever a student on the applicant file reports as dependent with a family size of four and two in college, best family size and best number in college were assigned using the probabilities given in the cells of the table. Similar distributions were determined and used for each combination of reported dependency status, family size, and number in college.

73

The cold-deck procedures described above are inappropriate for determining eligibility for applicants that report they are independent but who are, in fact, dependent. For such dependency status "switchers" it is necessary to impute all parental income data. The imputations must recreate a pattern of relationships between all imputed variables. To this end, for independent to dependent switchers, we employed a "hot-deck" imputation procedure.

In the hot-deck approach each switcher has a separately chosen set of family income variables imputed from among the "donor" values from dependent student applications. The hot-deck approach is currently in use in the Current Population Survey, Social Security Benefit Estimates, various Department of Energy Surveys, and is being tested on IRS Statistics of Income 1040 Series. Good theoretical discussions of hot-deck imputations can be found in Oh and Scheuren (1980), Welniah and Coder (1980), Chapman (1976) and Ernst (1980).

Hot-deck imputations were conducted using a two stage process. First, a probability of dependency status switch was calculated. For each applicant a switching status (yes or no) was assigned with a probability proportional to the switching rate.

Second, for each applicant assigned to a switching status a donor was selected from dependent applicants. The donor and recipient were matched by random selection with replacement. A similar approach was used for dependent to independent switchers.

IMPUTATION/ASSIGNMENT PROCEDURES

The accurate imputation of Stage III error to the applicant data base required systematic attention to numerous important details which occurred in three separate phases. First, analysis of the frequency, simultaniety, and level of error on the Stage III data base was necessary. Second, development of imputation software was required. Lastly, tests for goodness of fit were required to assess the accuracy of the imputation. Each of these phases is treated in the following sections of this appendix.

Analysis of Stage III Recipient Data

Data from the Stage III study were analyzed to determine the distribution of errors. This analysis involved three steps. The first step determined which cases had dependency status error. The second step determined which students had error in each variable. The third determined the degree of error for each variable.

Dependency Status Error. Dependency status error presented a unique problem and therefore was handled separately from all other errors. The following table summarizes the frequency of the two types of dependency status error found in the Stage III data.

Characteristics	Percentage of Cases with Dependency Status Error
• Students reporting as independent, unmarried, and living alone	16.9%
• Students reporting as independent and married or family size greater than one	8.5%
• Students reporting as dependent	.6%

These error rates were later imputed to the applicant file.

Cases selected as dependency switchers were handled differently than all other cases. Reported data and "best" data are unrelated for switchers. For example, students who report as independent report their own adjusted gross income. The "best" adjusted gross income for a student who switches to dependent is his parents' adjusted gross income which was not reported.

For each applicant selected as a switcher, a switcher (in the same direction) was randomly selected from the Stage III data base with replacement. The best values from the "donor" were then mapped onto the applicant record. No additional imputation procedures were required for dependency status switchers.

Presence or Absence of Error. For each variable, probability tables giving error rates conditioned on strata and zero/non-zero reported values were produced. These error rates were later used to impute error to applicants. As stated earlier, the presence or absence of error was assumed to be interdependent for some variables. Joint distributions of error were determined for these variables, again conditioned on strata and zero/non-zero reported values.

Degree of Error. For all but three of the eighteen variables, regression equations were determined to explain the degree of error. Student marital status was treated as a dichotomous variable (married/not married). Thus, if a case is determined to have an error in student marital status, the best value is the complement of the reported value.

Family size and number in college are discrete variables for which regression equations with sufficient prediction ability could not be determined. Instead, the joint distribution of best values for family size and number in college conditioned on respective reported values was determined. This joint distribution, given in Table C-1 of Appendix C, s later imputed to the applicant file.

Regression equations using Ordinary Least Squares (OLS) estimation were determined for each of the fifteen remaining variables within each stratum. Strata were collapsed for some variables to ensure sufficient degrees of freedom. For each variable, only Stage III cases with error in that variable were used in estimating the regression equations. The dependent variable in each regression was the computed best value. All explanatory variables were reported values or functions of reported values. In general, income and asset variables along with the reported value were used to explain the best values.

"Dummy" variables were used to explain the effects of zero reported values in the explanatory variables on best values. For each variable, a "dummy" variable was assigned. The "dummy" takes on the value zero when the variable it describes was zero, and a value of one otherwise.

Table C-2 of Appendix C lists the regression equations determined by OLS for each variable. Variables were stratified as shown in the table. Dependency status is given at the top of the page. The equations are grouped by dependent variable. Rows and columns represent income levels and explanatory variables, respectively. Each cell contains the OLS estimator for the regression coefficient for its respective income level and explanatory variable.

The column labeled "INTERCEPT" gives the OLS estimate of the best value when all other explanatory variables are zero. The column labeled "R-SQUARE" (R^2) gives a measure of how well the equation explains the variance in the dependent variable. R^2 is the ratio of variance explained by the regression equation to the total variance. An R^2 of one would indicate a perfect fit of the data to the equation. A zero R^2 would indicate that the equation explains none of the variance.

Imputation Software

The Statistical Analysis System (SAS) was used for all imputation software. The statistical procedures and file management capabilities of SAS were conducive to the imputation process.

Dependency Switchers. The first step in the production of software was to separate the Stage III data base into three separate files:

- Independent to dependent switchers
- Dependent to independent switchers
- Nonswitchers

A SAS program was written to compare reported dependency status to best dependency status for each Stage III Pell recipient and to place each case into the appropriate file. This program also used the SAS procedure "FREQ" to produce a table giving the rates of dependency status errors. These rates were then used to produce code to select switchers for the imputation of dependency status error.

The switcher program stratifies applicants into three groups using reported values: dependents, unmarried independent living alone, and all other independents. The program then generates a random number from a uniform distribution between 0 and 1 (U (0,1)) for each case. If this random number is less than or equal to the corresponding error rate, the case is selected as a switcher. If a case is not selected as a switcher the best dependency status is the reported dependency status. For switchers, the best dependency status is the complement of the reported dependency status.

The program then assigns best values to switchers. Switchers are divided into two groups: independent to dependent and dependent to independent. The Stage III records within each of the two switcher files are arbitrarily numbered 1 through n, where n is the number of cases in the file. A random integer J is generated from a U (1,n) distribution for each switcher. The applicant switcher is then assigned all best values from the Jth record on the appropriate Stage III switcher file. The imputation process is then complete for switchers.

Error Rates. Secondly the file containing nonswitchers was input to FREQ to produce tables of error rates for each variable. These rates were stratified by reported dependency status, income, and reported zero/not zero. The FREQ procedure also produced a disk file containing error rates for each variable within each stratum. The disk file of error rates was then input to a code generator (written in SAS) which produced the software to impute error rates.

The error rate imputation software determines to which stratum each case belongs and assigns the appropriate error rate for each variable. The program then generates a random number from a U (0,1) distribution. If the random number is less than or equal to the error rate the case is chosen to receive error. Otherwise, no error is assigned to the case for that variable. For each case not selected to receive error on a particular variable, the reported value is taken as the best value and the imputation process is complete for that variable within the case.

Best Values. The SAS procedure REG was used to obtain regression equations for each variable within each stratum. The REG procedure produced tables giving estimated regression coefficients and other statistics for each variable from the Stage III data base. Only those cases in error for a variable were used in determining the regression equation for that variable. The tables allowed us to make decisions about which strata (if any) to collapse to ensure sufficient degrees of freedom. After redefining the strata, REG was run again on the Stage III data. This iteration of REG produced both tables and a disk file containing regression coefficients for each variable within each stratum. The regression equations are given in Table C-2 of

Appendix C. The coefficients on the disk file were run through a code generator which produced the best value imputation software.

The best value imputation software assigned each applicant a regression equation for each variable for which the applicant was selected to have error. The equation assigned was dependent upon the applicant's stratum. The best value was then computed as the sum of the products of all regression coefficients with corresponding reported or "dummy" values. The concept of dummy variables was discussed earlier. For those cases not selected for error, the best value was set to the reported value.

Final Merge. The applicant switchers and nonswitchers with best values replacing reported values were merged onto one file using SAS. This new file was formatted identically to the original applicant file so as to be compatible with ED's applicant based model.

Software Validation

Several measures were taken to ensure quality in imputation software. All code was manually reviewed by the programmer and by other analysts. Code generators were used to reduce the probability of syntax errors. Code produced from generators was thoroughly checked. Imputation software was tested on Stage III data before using on applicant data base.

Testing of Dependency Status Software. The Stage III data base was treated as if it contained applicant data and was input to the dependency status software. The frequency of imputed dependency status error was then compared with the frequency of actual dependency status error. The best values mapped to the switchers were compared to the "donor" values. These measures ensured that the dependency status software was logically correct and produced imputed data stochastically consistent with the original dependency status data.

Testing of Error Rate Imputation Software. The Stage III file was again treated as if it contained applicant data to test the error rate imputation software. Imputed error rates were compared to actual error rates. The results confirmed that the imputation software yielded error rates consistent with actual error rates.

Testing of Best Value Software. Similarly the Stage III data was used to test the best value software. Mean imputed best values were compared by stratum to mean actual best values. Table C-3 of Appendix C displays the results of this comparison. These results confirm the validity of the best value software.

Testing of the Final Merge. To ensure that the final data tape created from the imputation process was compatible with ED's model extensive checks were performed. The imputed data base was compared to the original applicant data base record by record to verify that the two data sets were identically sorted. Fields containing variables not affected by the imputation process were compared between the original and the imputed data base. Ranges of all items on the imputed data base were compared to the ranges of respective items on the original file. Hexadecimal dumps from both files were compared. All of these tests ensured the compatibility of our data base to ED's model.

Imputation of Error to Applicant Data Base

The applicant data base was run through the programs described in the Imputation Software section. These programs replaced existing data items with imputed data. Dependency status error was assigned first. Cases selected as switchers received best values from Stage III "[illegible]ors" and were separated into a new file. Error rates were imputed next. Applicants were selected to have error at the rate of observed error in the Stage III data base for each variable. Best values were then assigned to these cases chosen to have error. Best values were computed by substituting reported values into regression equations obtained from Stage III data. Finally applicant switchers and nonswitchers were merged producing a file of imputed data.

Goodness of Fit Tests

The Stage III data base and the imputed applicant data base were compared to ensure that the distribution of error on the applicant file approximated the distribution of error on the Stage III file. Means of imputed and best values are displayed in Table C-3 of Appendix C. After having submitted our imputed data base to ED for recalculation of award, we found a savings of $215 million when error is eliminated from the applicant data base. This is comparable to the Pell QC Stage III study which estimated a savings of $220 million.

REFERENCES

Barr, R.S. and Turner, J.S. (1980) "Merging the 1977 Statistics of Income and the March 1978 Current Population Survey." Prepared for the Office of Tax Analysis, Washington, D.C., U.S. Department of the Treasury.

Brewer, K. R. (1979) "Discussion." Symposium on Incomplete Data: Preliminary Proceedings U.S. DHEW 219-224.

Chapman, D. W. (1976) "A Survey of Nonresponse Imputation Procedures." American Statistical Association, Proceedings of the Section on Social Statistics 245-251.

Cochran, W. G. (1977) Sampling Techniques. New York, John Wiley.

Ericson, W. A. (1969) "Subjective Bayesian Models in Sampling Finite Populations, I." Journal of the Royal Statistical Society Ser. 13, 31, 195-234.

Ernst, L. R. (1980) "Variance of the Mean for Several Hot-Deck Imputation Procedures." American Statistical Association, Proceedings of the Section on Survey Research Methods.

Oh, H. L. and Scheuren, F. J. (1980) "Estimating the Variance Impact of Missing CPS Income Data." American Statistical Association, Proceedings of the Section on Survey Research Methods.

Oh, H. L. and Scheuren F. J. (1981) "Weighting Adjustments for Unit Nonresponse." Panel on Incomplete Data, National Academy of Sciences.

Okner, B.A. (1972) "Constructing a New Data Base From Existing Microdata Sets: The 1966 Merge File." Annals of Economic and Social Measurement, 1, 325-342.

Radner, D.B. (1983). "Adjusted Estimates of the Size Distribution of Family Money Income." Journal of Business and Economic Statistics, 1, 136-146.

Radner, D.B., Allen, R., Gonzalez, M.E., Jabine, T.B., and Muller H.J. (1980) "Report on Exact and Statistical Matching Techniques." Statistical Policy Working Paper 5, U.S. Department of Commerce, Washington, D.C., U.S. Government Printing Office.

Rodgers, W.L. (1984) "An Evaluation of Statistical Matching." Journal of Business and Economic Statistics, Volume 2, No. 1, January 1984.

Rodgers, W.L. and DeVol, E. (1982). "An Evaluation of Statistical Matching." Unpublished report submitted to Income Survey Development Program. Department of Health and Human Services, Ann Arbor, Michigan, Institute for Social Research, The University of Michigan.

Royall, R. M. (1970) "On Finite Population Sampling Theory Under Certain Linear Regression Models." Biometrika, 57, 377-387

Schaible, W. L. (1979) "Estimation of Finite Population Totals from Incomplete Sample Data: Prediction Approaches." Symposium on Incomplete Data: Preliminary Proceedings DHEW 170-187.

TECHNICAL APPENDIX C

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

DEPENDENT STUDENTS

REPORTED FAMILY SIZE=2   REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	90.1974	91.9075	90.303
2	2	1.7145	2.8902	1.660
3	1	6.4016	4.6243	6.171
3	2	0.5649	0.0000	0.532
4	1	1.1218	0.5780	1.331

REPORTED FAMILY SIZE=2   REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	49.8291	70	46.527
2	2	40.6438	20	44.097
3	1	9.5272	10	9.375

REPORTED FAMILY SIZE=3   REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	11.1240	12.4567	10.797
2	2	0.7110	0.3460	0.588
3	1	78.7456	77.8567	79.272
3	2	1.3528	1.0381	1.205
4	1	3.5233	3.8062	3.630
4	2	0.7036	1.0381	0.786
5	1	1.7851	1.7301	1.631
6	1	0.6837	0.3460	0.720
6	2	0.3382	0.3460	0.426
8	1	0.3454	0.3460	0.332
9	4	0.3491	0.3460	0.272
10	1	0.3382	0.3460	0.308

REPORTED FAMILY SIZE=3   REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	14.7077	13.6842	14.615
2	2	6.3743	9.4737	6.531
3	1	5.4965	8.4211	5.160
3	2	64.9322	63.1579	65.932
3	3	2.0429	1.0526	2.337
4	2	2.2631	2.1053	1.660
4	3	2.1328	2.1053	1.760
5	1	1.0254	0.0000	1.118
5	2	1.0251	0.0000	0.974

REPORTED FAMILY SIZE=3   REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	2	68.4054	100.000	65.753
3	3	31.5946	0.000	34.24

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

REPORTED FAMILY SIZE=4 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	[illegible]	1.0709	1.7857	0.9332
2	[illegible]	0.7327	0.7143	0.6978
3	1	15.9934	16.7857	16.1917
3	2	0.7183	0.7143	0.6810
4	1	71.4897	70.0000	71.0635
4	2	4.2050	2.8571	4.3716
5	1	4.3017	6.0714	4.4725
5	2	0.7039	0.7143	0.8911
6	1	0.7845	0.3571	0.6978

REPORTED FAMILY SIZE=4 REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	2.4882	1.1905	2.4255
2	2	1.2899	2.3810	1.3444
3	1	13.5541	15.4762	13.2937
3	2	4.6344	3.5714	4.8517
3	3	1.2446	0.5952	1.1106
4	1	8.2185	7.1429	8.6073
4	2	59.4057	58.9286	59.7399
4	3	0.5890	0.0000	0.6284
5	1	0.6387	1.1905	0.6138
5	2	6.1334	7.1429	5.8746
6	2	0.5859	1.1905	0.5115
6	3	0.5550	0.5952	0.5553
7	3	0.6317	0.5952	0.6430

REPORTED FAMILY SIZE=4 REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	6.0671	2.9412	4.5400
2	2	3.0032	8.8235	2.9869
3	1	2.9711	0.0000	2.1505
3	2	23.2671	26.4706	22.3417
3	3	8.9610	2.9412	10.6358
4	1	2.7191	2.9412	4.1816
4	2	6.0312	8.8235	7.1685
4	3	41.1327	38.2353	39.6655
5	3	2.7887	8.8235	3.4648
7	1	3.0590	0.0000	3.4648

REPORTED FAMILY SIZE=4 REPORTED # IN COLLEGE=4

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
4	4	100.000	100	100

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

DEPENDENT STUDENTS

REPORTED FAMILY SIZE=5   REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	1.6694	0.5587	1.5826
3	1	7.8537	6.7039	7.9777
4	1	18.4696	22.9050	18.9648
4	2	1.6616	1.1173	1.7642
5	1	65.2999	63.6872	64.9111
5	2	2.7821	4.4693	2.8279
6	1	1.1479	0.0000	0.9729
6	2	1.1159	0.5587	0.9988

REPORTED FAMILY SIZE=5   REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	4.4949	2.5974	3.7819
3	2	0.6437	1.2987	0.8104
4	1	10.8049	12.3377	11.3625
4	2	13.6017	11.0390	14.1989
5	1	6.5426	5.8442	6.6659
5	2	54.6289	56.4935	54.2630
5	3	4.6775	6.4935	4.5079
6	1	1.3176	2.5974	1.2156
6	2	1.3697	0.6494	1.4351
6	3	1.9184	0.6494	1.7559

REPORTED FAMILY SIZE=5   REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	3.6209	3.5714	3.1937
3	1	3.4568	1.7857	4.0858
3	2	3.5448	3.5714	3.7696
4	1	3.3989	8.9286	3.4555
4	2	13.9051	19.6429	13.1937
4	3	1.6361	1.7857	1.5183
4	4	1.7509	0.0000	2.1466
5	1	3.3812	1.7857	3.8743
5	2	14.2421	16.0714	14.8168
5	3	43.9519	33.9286	43.3508
5	4	1.7628	3.5714	1.5183
6	3	3.5527	3.5714	3.5079
7	1	1.7877	1.7857	1.5707

REPORTED FAMILY SIZE=5   REPORTED # IN COLLEGE=4

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
4	3	12.0957	0.0000	9.2683
5	2	20.5168	55.5556	29.2683
5	3	53.7081	11.1111	31.7073
5	4	11.2094	0.0000	13.1707
5	5	11.0524	0.0000	9.2683
6	4	11.4165	33.3333	7.3171

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

DEPENDENT STUDENTS

REPORTED FAMILY SIZE=6 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	2.0727	2.1277	2.5941
3	1	2.1683	0.0000	2.2380
4	1	11.1845	8.5106	10.7833
5	1	18.7997	13.0851	19.9898
5	3	1.0518	1.0638	0.9664
6	1	55.1362	60.6383	54.6796
6	2	2.1677	3.1915	1.7803
7	1	4.1907	3.1915	3.7640
7	2	1.0518	2.1277	1.0173
9	1	1.0356	1.0638	1.1953
10	1	1.1409	0.0000	0.9919

REPORTED FAMILY SIZE=6 REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	2.2062	5.3763	2.4253
3	2	1.1209	2.1505	0.9588
4	1	3.1201	4.3011	3.2713
4	2	5.2811	3.2258	4.7941
5	1	11.5371	11.8280	11.5623
5	2	12.0895	7.5269	11.7315
6	1	6.2054	5.3763	6.4016
6	2	52.7071	53.7634	53.6379
6	3	3.3625	5.3763	3.2995
6	4	1.2492	0.0000	0.9306
7	2	1.1209	1.0753	0.9870

REPORTED FAMILY SIZE=6 REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	1.6926	0.0000	1.8892
3	2	1.6926	0.0000	1.7632
4	1	3.4657	5.4545	3.4635
4	3	1.7732	0.0000	1.8892
5	1	1.8085	0.0000	2.3300
5	2	16.2950	18.1818	15.9950
5	3	9.3706	10.9091	9.0680
6	1	1.6926	1.8182	1.8892
6	2	5.4623	3.6364	4.8489
6	3	47.6870	49.0909	47.4181
6	4	1.8591	1.8182	2.3300
8	2	5.4948	9.0909	5.1008
9	3	1.73591	0	2.01511

REPORTED FAMILY SIZE=6 REPORTED # IN COLLEGE=4

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
5	1	6.6317	0.0000	7.2289
5	2	6.6317	14.2857	5.7229
5	3	13.9471	21.4286	18.3735
5	5	7.1191	7.1429	6.6265
6	2	6.6317	0.0000	5.7229
6	3	15.7198	7.1429	13.5542
6	4	43.3189	50.0000	42.7711

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

DEPENDENT STUDENTS

REPORTED FAMILY SIZE=7 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	5.8490	2.7778	7.0232
4	1	8.2094	11.1111	8.7520
5	1	8.2334	2.7778	7.5635
6	1	13.6009	16.6667	14.1005
6	2	2.6747	16.6667	3.1875
7	1	36.4864	33.3333	34.3058
7	2	16.4611	8.3333	15.9373
8	1	8.4850	8.3333	9.1302

REPORTED FAMILY SIZE=7 REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	2.1475	4.0816	1.5986
4	1	2.1107	2.0408	2.3091
5	1	6.0453	8.1633	6.6311
5	2	8.2793	4.0816	7.7561
6	1	15.4872	8.1633	15.8674
6	2	2.0047	2.0408	2.2499
6	3	2.0347	0.0000	1.9538
7	1	1.8762	0.0000	1.5986
7	2	53.6732	65.3061	54.2333
9	3	2.1890	2.0408	2.0722
10	2	2.1475	2.0408	1.7762
10	3	2.0047	2.0408	1.9538

REPORTED FAMILY SIZE=7 REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
5	1	2.4217	2.6316	2.3230
5	2	5.0094	7.8947	5.4204
5	3	2.6468	0.0000	3.3186
6	2	15.4406	10.5263	14.4912
6	3	10.7545	18.4211	8.8496
7	2	8.6064	15.7895	8.0752
7	3	49.3523	39.4737	51.5487
7	5	2.6952	2.6316	2.4336
8	4	3.0737	2.6316	3.5398

87

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

DEPENDENT STUDENTS

------------------------ REPORTED FAMILY SIZE=7 REPORTED # IN COLLEGE=4 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
6	3	39.2919	60	39.8601
7	3	20.3936	0	18.8811
7	4	40.3145	40	41.2587

------------------------ REPORTED FAMILY SIZE=7 REPORTED # IN COLLEGE=5 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
7	4	100.000	100	100.000

------------------------ REPORTED FAMILY SIZE=7 REPORTED # IN COLLEGE=6 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	100.000	100	100.000

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

DEPENDENT STUDENTS

REPORTED FAMILY SIZE=8 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
6	1	5.6260	0.0000	5.9250
7	1	21.3962	21.0526	24.1838
8	1	36.6746	15.7895	34.9156
8	2	10.3131	21.0526	11.0036
9	1	15.6198	31.5789	15.1149
9	2	4.9456	5.2632	4.1112
10	1	5.4249	5.2632	4.7158

REPORTED FAMILY SIZE=8 REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
4	1	3.8976	0	3.5669
7	1	11.9956	16	12.6115
7	2	20.0932	44	19.7452
7	3	3.9651	0	6.0764
8	1	4.2088	0	3.8217
8	2	35.9299	28	37.9618
8	3	3.9859	4	2.6752
9	1	12.0201	4	11.8471
9	2	3.9038	4	3.6943

REPORTED FAMILY SIZE=8 REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	6.4890	0.0000	7.9051
5	1	6.9446	7.1429	6.1265
5	3	6.9919	14.2857	7.1146
6	3	6.9919	0.0000	5.5336
7	2	13.5997	21.4286	12.6482
7	3	11.4846	14.2857	22.1344
8	3	37.4982	42.8571	38.5375

REPORTED FAMILY SIZE=8 REPORTED # IN COLLEGE=4

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
4	2	9.5926	0	10.3448
5	1	19.1623	40	22.4138
6	1	9.7996	30	11.3943
7	3	10.1906	0	14.9425
8	3	20.2483	10	10.3448
8	4	31.0066	20	30.4598

REPORTED FAMILY SIZE=8 REPORTED # IN COLLEGE=5

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
7	4	100.000	100	100.000

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

DEPENDENT STUDENTS

------------------------ REPORTED FAMILY SIZE=9 REPORTED # IN COLLEGE=1 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	7.4066	7.6923	8.5393
4	1	7.4336	7.6923	8.7640
5	1	8.3049	7.6923	7.4157
6	1	7.6440	7.6923	7.6404
7	1	30.0630	38.4615	32.5843
6	2	8.3049	7.6923	6.5169
9	1	30.8430	23.0769	28.5393

------------------------ REPORTED FAMILY SIZE=9 REPORTED # IN COLLEGE=2 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
5	1	14.0232	14.2857	12.5356
6	1	14.1651	14.2857	11.3960
9	1	14.1651	28.5714	14.5299
9	2	57.6465	42.8571	61.5385

------------------------ REPORTED FAMILY SIZE=9 REPORTED # IN COLLEGE=3 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
7	1	18.7804	0	19.2771
7	3	20.3673	60	21.2851
8	3	18.7804	0	22.4900
9	3	21.9111	20	17.2691
9	4	20.1608	20	19.6787

------------------------ REPORTED FAMILY SIZE=9 REPORTED # IN COLLEGE=4 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
8	2	24.9702	50	27.5862
9	3	24.0080	0	25.2874
9	4	51.0218	50	47.1264

------------------------ REPORTED FAMILY SIZE=9 REPORTED # IN COLLEGE=5 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
9	5	100.000	100	100.000

90

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

DEPENDENT STUDENTS

REPORTED FAMILY SIZE=10 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
9	1	53.3713	0	49.5575
10	1	46.6287	100	50.4425

REPORTED FAMILY SIZE=10 REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	33.5949	0.000	24.2424
10	2	66.4051	100.000	75.7576

REPORTED FAMILY SIZE=10 REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
9	1	25.1126	25	21.2389
10	3	74.8874	75	78.7611

REPORTED FAMILY SIZE=10 REPORTED # IN COLLEGE=4

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
10	4	100.000	100	100.000

REPORTED FAMILY SIZE=10 REPORTED # IN COLLEGE=6

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
10	6	100.000	100	100.000

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

DEPENDENT STUDENTS

REPORTED FAMILY SIZE=11 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
5	1	32.5810	66.6667	34.3750
9	1	35.3415	33.3333	41.6667
11		32.0775	0.0000	23.9583

REPORTED FAMILY SIZE=11 REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
10	2	32.9473	0.000	33.3333
11	2	67.0527	100.000	66.6667

REPORTED FAMILY SIZE=11 REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
9	2	48.1480	100	50.0000
11	3	51.8520	0	50.0000

REPORTED FAMILY SIZE=12 REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
10	2	100.000	100	100.000

REPORTED FAMILY SIZE=13 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
11	1	100.000	100	100.000

REPORTED FAMILY SIZE=14 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
7	1	100.000	100	100.000

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

INDEPENDENT STUDENTS

------------------------ REPORTED FAMILY SIZE=1 REPORTED # IN COLLEGE=1 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
1	1	96.2365	95.9335	96.5089
2	1	1.5129	2.2181	1.6828
2	2	0.2335	0.1848	0.2072
3	1	0.6163	0.9242	0.6772
4	1	0.2335	0.0000	0.1758
5	1	0.7004	0.1848	0.5777
5	2	0.2335	0.1848	0.2260
7	1	0.2335	0.3697	0.1444

------------------------ REPORTED FAMILY SIZE=2 REPORTED # IN COLLEGE=1 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
1	1	6.7887	8.7912	6.4718
2	1	83.6236	83.5165	84.1251
2	2	4.[illegible]389	1.6484	4.0114
3	1	4.7763	5.4945	4.8640
3	2	0.5725	0.5495	0.5278

------------------------ REPORTED FAMILY SIZE=2 REPORTED # IN COLLEGE=2 ------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
1	1	13.0065	12.7273	11.4331
2	1	19.8568	16.3636	19.1811
2	2	63.7023	67.2727	66.1417
3	2	3.4343	3.6364	3.2441

93

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

INDEPENDENT STUDENTS

REPORTED FAMILY SIZE=3 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
1	1	2.1775	2.8777	2.0662
2	1	1.5056	0.0000	1.3390
3	1	91.2817	92.8058	91.3541
3	2	1.6401	1.4388	1.4809
4	1	2.7638	2.8777	3.0416
5	1	0.6314	0.0000	0.7183

REPORTED FAMILY SIZE=3 REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	1	3.4331	4	3.9805
2	2	8.9075	12	8.6921
3	1	24.2904	28	23.8018
3	2	52.3596	48	50.8123
4	1	3.7015	0	4.5898
4	2	7.3080	8	8.1235

REPORTED FAMILY SIZE=3 REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
1	1	15.8275	16.6667	9.3750
3	2	17.8914	33.3333	18.7500
3	3	49.5244	50.0000	59.3750
4	2	16.7567	0.0000	12.5000

94

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

INDEPENDENT STUDENTS

REPORTED FAMILY SIZE=4 REPORTED # IN COLLEGE=1

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
1	1	2.1556	2.2472	2.3381
2	1	1.0728	0.0000	1.0261
3	1	4.6498	4.4944	4.7183
4	1	83.3795	78.6517	82.4054
4	2	5.4491	11.2360	5.9294
5	1	2.1508	3.3708	2.5147
6	1	1.1423	0.0000	1.0681

REPORTED FAMILY SIZE=4 REPORTED # IN COLLEGE=2

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
1	1	6.7351	0.0000	7.4268
2	2	6.7534	6.6667	6.5202
3	1	18.9664	13.3333	20.0488
3	2	7.9857	20.0000	6.6248
4	1	19.9505	6.6667	20.3626
4	2	32.8738	46.6667	32.1827
5	2	6.7351	6.6667	6.8340

REPORTED FAMILY SIZE=4 REPORTED # IN COLLEGE=3

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
2	2	21.7344	0	22.3451
4	3	78.2656	100	77.6549

REPORTED FAMILY SIZE=4 REPORTED # IN COLLEGE=4

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
4	4	100.000	100	100.000

95

BEST COPY AVAILABLE

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

INDEPENDENT STUDENTS

------------------------- REPORTED FAMILY SIZE=5 REPORTED # IN COLLEGE=1 -------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	6.1955	6.2500	6.0332
4	1	6.4959	15.6250	5.7483
5	1	74.8820	65.6250	75.2304
6	1	12.4265	12.5000	12.9881

------------------------- REPORTED FAMILY SIZE=5 REPORTED # IN COLLEGE=2 -------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
3	1	12.5583	44.4444	11.7506
4	1	22.6351	11.1111	22.2422
4	2	10.0808	22.2222	10.9113
5	2	54.7298	22.2222	55.0959

------------------------- REPORTED FAMILY SIZE=5 REPORTED # IN COLLEGE=3 -------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
4	1	16.0854	16.6667	15.8287
4	2	15.2194	0.0000	16.9460
5	3	33.0595	16.6667	35.3818
5	4	18.6883	33.3333	17.1322
6	2	16.9874	33.3333	14.7114

------------------------- REPORTED FAMILY SIZE=5 REPORTED # IN COLLEGE=4 -------------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
5	3	100.000	100	100.000

TABLE C-1

JOINT DISTRIBUTION OF "BEST" FAMILY SIZE AND "BEST" NUMBER IN COLLEGE BY RESPECTIVE REPORTED VALUES

INDEPENDENT STUDENTS

-------------------- REPORTED FAMILY SIZE=6 REPORTED # IN COLLEGE=1 --------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
5	1	6.6912	23.0769	8.1327
6	1	84.8463	76.9231	84.0981
6	2	8.4624	0.0000	7.7692

-------------------- REPORTED FAMILY SIZE=6 REPORTED # IN COLLEGE=2 --------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
6	1	34.6375	33.3333	31.2907
6	2	65.3625	66.6667	68.7093

-------------------- REPORTED FAMILY SIZE=7 REPORTED # IN COLLEGE=1 --------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
6	1	20.9688	40	22.9885
7	1	79.0312	60	77.0115

-------------------- REPORTED FAMILY SIZE=7 REPORTED # IN COLLEGE=2 --------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
7	2	100.000	100	100.000

-------------------- REPORTED FAMILY SIZE=8 REPORTED # IN COLLEGE=1 --------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
6	1	47.1514	50	48.1481
8	1	52.8486	50	51.8519

-------------------- REPORTED FAMILY SIZE=8 REPORTED # IN COLLEGE=3 --------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
5	2	48.5741	50	44.4444
8	3	51.4259	50	55.5556

-------------------- REPORTED FAMILY SIZE=9 REPORTED # IN COLLEGE=2 --------------------

'BEST' FAMILY SIZE	'BEST' # IN COLLEGE	STAGE III ACTUAL PERCENT	STAGE III IMPUTED PERCENT	APPLICANT IMPUTED PERCENT
		100.000	100	100.000

97

TABLE C-2

REGRESSION EQUATIONS USED FOR IMPUTATION OF BEST VALUES TO APPLICANT DATA BASE

Dependent Students

DEPENDENT VARIABLE=ADJUSTED GROSS INCOME

INCOME LEVEL	AGI	DUMMY (AGI)	NONTAX INCOME	TAXES PAID	DUMMY (NONTAX INC)	DUMMY (TAX)
UP TO $8,000	0.783123	-3634	0.24625	0.58976	-1582.4	-1565.0
$8,000 TO $15,000	0.942472	-15473	-0.10601	-0.57325	-203.7	-614.4
$15,000 TO $20,000	0.888809	-14119	-0.81855	0.23109	366.8	877.5
OVER $20,000	0.975143	0	-0.06064	-0.00206	-218.0	-962.4

INCOME LEVEL	NET HOME VALUE	DUMMY (NET HOME VALUE)	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $8,000	0.00378	-262.44	0.0032219	-260.28	5382.5	0.2074
$8,000 TO $15,000	-0.15032	2108.11	0.0844767	49.70	16181.5	0.2509
$15,000 TO $20,000	-0.06315	1329.12	0.0006868	842.33	14161.7	0.6508
OVER $20,000	-0.02938	526.75	0.0096555	-52.09	1756.2	0.7808

DEPENDENT VARIABLE=NONTAXABLE INCOME

INCOME LEVEL	AGI	DUMMY (AGI)	NONTAX INCOME	FAMILY SIZE
UP TO $8,000	-0.088025	-150.1	0.77754	36.035
$8,000 TO $15,000	0.049151	-2040.7	0.83839	111.488
$15,000 TO $20,000	0.100195	-4423.0	1.23457	46.636
OVER $20,000	-0.105731	0.0	0.87486	8.648

INCOME LEVEL	DUMMY (NONTAX INC)	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $8,000	-419.35	-0.012429	-282.76	2426.53	0.1049
$8,000 TO $15,000	757.54	0.009551	-446.25	1577.34	0.4462
$15,000 TO $20,000	185.72	-0.005204	1241.21	501.73	0.4029
OVER $20,000	505.11	-0.002381	37.40	489.73	.4132

DEPENDENT VARIABLE=NET HOME VALUE

INCOME LEVEL	AGI	DUMMY (AGI)	NET HOME VALUE	DUMMY (NET HOME VALUE)
UP TO $8,000	-0.21411	2135	0.226854	-16384
$8,000 TO $15,000	0.24789	-7525	0.715202	-29854
$15,000 TO $20,000	-0.20023	11798	1.010628	-19125
OVER $20,000	0.18202	-37445	0.933369	-21112

INCOME LEVEL	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $8,000	0.28042	11481.8	13308.4	0.1728
$8,000 TO $15,000	0.12252	7059.7	19138.6	0.3836
$15,000 TO $20,000	-0.12227	14105.7	1414.9	0.4178
OVER $20,000	-0.11395	5730.6	54598.9	0.3417

TABLE C-2

REGRESSION EQUATIONS USED FOR IMPUTATION OF BEST VALUES TO APPLICANT DATA BASE

Dependent Students

DEPENDENT VARIABLE=TAXES PAID

INCOME LEVEL	AGI	DUMMY (AGI)	TAXES PAID	DUMMY (TAX)	NONTAX INCOME	DUMMY (NONTAX INC)
UP TO $8,000	-0.003003	-86.1	0.233975	-409.6	-0.069441	107.64
$8,000 TO $15,000	0.090397	-617.1	-0.002217	-317.4	0.112340	-158.00
$15,000 TO $20,000	0.109911	-1974.2	0.096413	-772.6	0.050771	-24.89
OVER $20,000	0.037072	0.0	0.093484	-2598.5	-0.096055	451.19

INCOME LEVEL	NET HOME VALUE	DUMMY (NET HOME VALUE)	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $8,000	-0.005404	-108.00	0.0075227	7.49	508.82	0.3858
$8,000 TO $15,000	-0.001779	-3.41	-0.0047088	128.86	586.28	0.2332
$15,000 TO $20,000	-0.017758	356.88	0.0031147	-147.86	1349.32	0.5016
OVER $20,000	0.012032	-737.83	0.0038328	1156.23	1249.98	0.4564

DEPENDENT VARIABLE=FATHER PORTION

INCOME LEVEL	FATHER PORTION	AGI	DUMMY (AGI)	NET ASSETS
UP TO $8,000	0.912025	0.540042	-9721.2	-0.012142
$8,000 TO $15,000	0.674069	0.439376	241.2	-0.016158
$15,000 TO $20,000	0.751206	0.579111	-5151.0	-0.007981
OVER $20,000	0.768886	0.319702	0.0	0.002763

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (STUD. PORT.)	INTERCEPT	R-SQUARE
UP TO $8,000	419.0	-4633	12529.1	0.1107
$8,000 TO $15,000	-1058.0	-4112	4105.7	0.4380
$15,000 TO $20,000	1379.0	-10121	7905.7	0.4852
OVER $20,000	1244.7	-15621	11511.7	0.6446

DEPENDENT VARIABLE=MOTHER PORTION

INCOME LEVEL	MOTHER PORTION	AGI	DUMMY (AGI)	NET ASSETS
UP TO $8,000	0.068883	0.52755	-3922.7	0.013503
$8,000 TO $15,000	0.413092	1.39508	-4890.5	-0.021775
$15,000 TO $20,000	0.495107	-0.01670	-502.3	-0.015712
OVER $20,000	0.399047	-0.12577	0.0	-0.062372

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (MOTHER PORT.)	INTERCEPT	R-SQUARE
UP TO $8,000	-905.41	1362.7	4303.0	0.2534
$8,000 TO $15,000	415.40	-7227.3	8978.3	0.5530
$15,000 TO $20,000	1036.80	-7716.9	7120.5	0.7758
OVER $20,000	-362.30	-3925.1	13337.7	0.3763

TABLE C-2

REGRESSION EQUATIONS USED FOR IMPUTATION OF BEST VALUES TO APPLICANT DATA BASE

Dependent Students

DEPENDENT VARIABLE=DEPENDENT'S INCOME

INCOME LEVEL	AGI	DUMMY (AGI)	DEPENDENT'S INCOME	DEPENDENT'S ASSETS	DUMMY (DEP. INCOME)
UP TO $8,000	0.022330	-616.5	0.094106	0.08631	-611.5
$8,000 TO $15,000	-0.019806	-51.8	0.848334	0.34749	-1713.9
$15,000 TO $20,000	0.035811	1505.1	0.319345	0.19435	-694.7
OVER $20,000	-0.002840	-1412.5	0.969050	-0.19803	-1903.7

INCOME LEVEL	DUMMY (DEPENDENT'S ASSETS)	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $8,000	211.44	-0.0037001	115.13	2266.74	0.1724
$8,000 TO $15,000	22.92	-0.0009476	-204.87	2493.50	0.5308
$15,000 TO $20,000	-439.91	0.0042141	-401.88	589.03	0.1976
OVER $20,000	9.51	-0.0031206	541.16	3128.69	0.5791

DEPENDENT VARIABLE=DEPENDENT'S ASSETS

INCOME LEVEL	AGI	DUMMY (AGI)	DEPENDENT'S ASSETS	DEPENDENT'S INCOME	DUMMY (DEPENDENT'S ASSETS)
UP TO $8,000	-0.016793	-163.88	0.38392	-0.006191	-178.81
$8,000 TO $15,000	-0.034964	37.91	0.45605	-0.006895	-16.73
$15,000 TO $20,000	-0.030827	-379.96	0.31948	-0.010545	-36.31
OVER $20,000	0.012778	335.63	1.00594	-0.039568	-188.25

INCOME LEVEL	DUMMY (DEP. INCOME)	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $8,000	-262.64	0.0032831	200.096	457.92	0.0305
$8,000 TO $15,000	57.33	0.0038600	132.721	63.38	0.1242
$15,000 TO $20,000	16.67	0.0012737	33.488	534.27	0.0842
OVER $20,000	93.64	-0.0016026	-85.418	-302.90	0.1792

DEPENDENT VARIABLE=TUITION

INCOME LEVEL	TUITION	AGI	DUMMY (AGI)	NET ASSETS
ALL LEVELS	0.932327	0.00401759	-78.886	0.0020668

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (TUITION)	INTERCEPT	R-SQUARE
ALL LEVELS	156.962	-506.45	382.435	0.8338

1..0

TABLE C-2

REGRESSION EQUATIONS USED FOR IMPUTATION OF BEST VALUES TO APPLICANT DATA BASE

Dependent Students

DEPENDENT VARIABLE=MEDICAL/DENTAL

INCOME LEVEL	MEDICAL DENTAL	AGI	DUMMY (AGI)	NET ASSETS
UP TO $8,000	0.649448	-0.01396	225.77	-0.0003102
$8,000 TO $15,000	0.651055	0.014330	-358.42	-0.0016306
$15,000 TO $20,000	0.646145	0.011662	183.60	-0.0015787
OVER $20,000	0.759400	-0.000489	0.00	-0.0005477

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (MEDICAL)	INTERCEPT	R-SQUARE
UP TO $8,000	-203.02	-295.97	633.35	0.4620
$8,000 TO $15,000	36.01	-876.57	996.84	0.3065
$15,000 TO $20,000	498.82	-210.35	-307.72	0.4484
OVER $20,000	117.29	-567.25	728.33	0.4984

DEPENDENT VARIABLE=EDUCATIONAL SOCIAL SECURITY

INCOME LEVEL	EDUCATIONAL SOC. SEC.	AGI	DUMMY (AGI)	NET ASSETS
UP TO $8,000	0.807735	0.0168211	-237.28	-0.0038287
$8,000 TO $15,000	0.203552	0.0073822	-210.81	0.0040680
OVER $15,000	0.257026	0.0254220	88.13	0.0071256

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (SS)	INTERCEPT	R-SQUARE
UP TO $8,000	-308.59	-931.34	1237.38	0.4279
$8,000 TO $15,000	-85.03	-302.67	1188.44	0.1242
OVER $15,000	-583.25	-672.49	1125.13	0.3189

DEPENDENT VARIABLE=EDUCATIONAL VA

INCOME LEVEL	EDUCATIONAL VA BENEFITS	INTERCEPT	R-SQUARE
ALL LEVELS	-0.23419	2266.26	0.0000

101

TABLE C-2

REGRESSION EQUATIONS USED FOR IMPUTATION OF BEST VALUES TO APPLICANT DATA BASE

Dependent Students

DEPENDENT VARIABLE=CASH AND ACCOUNTS

INCOME LEVEL	CASH AND ACCOUNTS	AGI	DUMMY (AGI)	NET ASSETS
UP TO $8,000	0.261597	0.0270536	249.32	0.0119430
$8,000 TO $15,000	0.522089	0.0291546	259.03	0.0027900
$15,000 TO $20,000	1.003559	0.0408384	0.00	0.0323474
OVER $20,000	0.379011	-0.0035430	-498.01	0.0236792

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (CASH)	INTERCEPT	R-SQUARE
UP TO $8,000	329.23	673.457	-24.91	0.2565
$8,000 TO $15,000	614.09	-60.987	-329.34	0.4095
$15,000 TO $20,000	-691.21	424.678	-511.03	0.4901
OVER $20,000	-426.39	605.382	982.92	0.2191

DEPENDENT VARIABLE=NET INVESTMENT VALUE

INCOME LEVEL	NET INVESTMENT VALUE	AGI	DUMMY (AGI)	NET ASSETS
UP TO $15,000	0.80491	-0.53067	14974.6	0.107138
$15,000 TO $20,000	1.04099	0.32327	0.0	-0.090237
OVER $20,000	0.94173	0.12302	-2050.2	0.057538

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (INVESTMENT)	INTERCEPT	R-SQUARE
UP TO $15,000	-3436.0	-6259	-2896.3	0.5129
$15,000 TO $20,000	0.0	-7520	6217.0	0.6647
OVER $20,000	0.	-12770	11475.6	0.5967

DEPENDENT VARIABLE=NET BUSINESS FARM VALUE

INCOME LEVEL	NET BUSINESS FARM VALUE	AGI	DUMMY (AGI)	NET ASSETS
UP TO $15,000	0.96633	-2.0675	-3585.5	-0.10631
OVER $15,000	1.18449	0.2723	33855.0	-0.21979

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (BUS. FARM VALUE)	INTERCEPT	R-SQUARE
UP TO $15,000	0	-38004	66739	0.6022
OVER $15,000	0	-17218	-20630	0.5664

102

BEST COPY AVAILABLE

TABLE C-2

REGRESSION EQUATIONS USED FOR IMPUTATION OF BEST VALUES TO APPLICANT DATA BASE

Independent Students

DEPENDENT VARIABLE=ADJUSTED GROSS INCOME

INCOME LEVEL	AGI	DUMMY (AGI)	NONTAX INCOME	TAXES PAID	DUMMY (NONTAX INC)
UP TO $2,000	0.48703	-3246.2	2.13028	4.22192	-1917.6
$2,000 TO $4,000	-0.05416	1433.8	-0.19969	1.52232	-738.3
$4,000 TO $8,000	0.71372	-1240.3	0.38562	1.41796	-612.9
OVER $8,000	1.09132	-5612.9	0.15910	0.23149	-600.4

INCOME LEVEL	DUMMY (TAX)	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $2,000	-774.55	0.457699	-916.65	4130.62	0.2217
$2,000 TO $4,000	1314.97	0.877781	-537.69	1202.64	0.4148
$4,000 TO $8,000	352.04	0.213757	-322.26	2075.18	0.3506
OVER $8,000	-489.40	-0.024398	-246.47	2736.00	0.8982

DEPENDENT VARIABLE=NONTAXABLE INCOME

INCOME LEVEL	AGI	DUMMY (AGI)	NONTAX INCOME	FAMILY SIZE
UP TO $2,000	0.02968	-703.5	0.75950	185.39
$2,000 TO $4,000	-0.15373	-419.5	1.96091	-314.63
$4,000 TO $8,000	-0.19277	226.5	1.72160	151.05
OVER $8,000	-0.05381	-3128.6	1.11460	-102.50

INCOME LEVEL	DUMMY (NONTAX INC)	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $2,000	-93.51	0.0376948	-583.09	1260.86	0.2639
$2,000 TO $4,000	56.00	0.0031256	66.33	1763.90	0.7391
$4,000 TO $8,000	324.71	0.1031900	-192.57	1223.11	0.7458
OVER $8,000	[illegible]	-0.0074209	-319.13	5222.85	0.7083

DEPENDENT VARIABLE=NET HOME VALUE

INCOME LEVEL	AGI	DUMMY (AGI)	NET HOME VALUE	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $4,000	-4.0311	29533.3	1.96166	-0.93678	-5024.6	5034.03	0.4101
$4,000 TO $8,000	-1.5556	3152.3	1.06395	-1.06319	-5470.8	6900.70	0.8257
OVER $8,000	0.3048	295.0	1.39858	0.18165	-1380.7	2011.05	0.2207

BEST COPY AVAILABLE

TABLE C-2

REGRESSION EQUATIONS USED FOR IMPUTATION OF BEST VALUES TO APPLICANT DATA BASE

Independent Students

------ DEPENDENT VARIABLE=TAXES PAID ------

INCOME LEVEL	AGI	DUMMY (AGI)	TAXES PAID	DUMMY (TAX)
UP TO $4,000	0.0312059	-40.57	-0.008519	-436.04
OVER $4,000	0.0319722	-327.48	0.309946	-41.60

INCOME LEVEL	NET ASSETS	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $4,000	0.281219	-111.42	452.286	0.3570
OVER $4,000	0.019244	59.08	347.919	0.3447

------ DEPENDENT VARIABLE=STUDENT PORTION ------

INCOME LEVEL	STUDENT PORTION	AGI	DUMMY (AGI)	NET ASSETS
UP TO $2,000	0.228480	0.652380	-1883.9	0.131227
$2,000 TO $4,000	0.569942	0.123951	1355.3	1.468630
$4,000 TO $8,000	0.591632	0.207977	123.3	0.134177
OVER $8,000	0.827929	0.077727	0.0	-0.000187

INCOME LEVEL	DUMMY (STUD. PORT.)	DUMMY (ASSETS)	INTERCEPT	R-SQUARE
UP TO $2,000	-1641.1	-231.30	4125.60	0.1621
$2,000 TO $4,000	-2406.5	27.67	1898.35	0.3143
$4,000 TO $8,000	-2229.9	11.43	3123.03	0.3453
OVER $8,000	-6442.3	1157.28	6774.88	0.6272

------ DEPENDENT VARIABLE=SPOUSE PORTION ------

INCOME LEVEL	SPOUSE PORTION	AGI	DUMMY (AGI)	NET ASSETS
ALL LEVELS	0.890035	0.158880	2480.54	0.0495138

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (SPOUSE PORT.)	INTERCEPT	R-SQUARE
ALL LEVELS	92.5736	-6463.6	3006.64	0.7618

104

------ DEPENDENT VARIABLE=TUITION ------

INCOME LEVEL	TUITION	INTERCEPT	R-SQUARE
ALL LEVELS	1.18859	28.3816	0.8211

TABLE C-2

REGRESSION EQUATIONS USED FOR IMPUTATION OF BEST VALUES TO APPLICANT DATA BASE

Independent Students

DEPENDENT VARIABLE-MEDICAL/DENTAL

INCOME LEVEL	MEDICAL DENTAL	AGI	DUMMY (AGI)	NET ASSETS
UP TO $4,000	[illegible]	-0.033431	-82.928	0.0087867
OVER $4,000	0.214396	-0.005048	0.000	0.0058458

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (MEDICAL)	INTERCEPT	R-SQUARE
UP TO $4,000	[illegible]	-115.47	303.838	0.6153
OVER $4,000	49.2416	-208.34	328.204	0.8546

DEPENDENT VARIABLE-CASH AND ACCOUNTS

INCOME LEVEL	CASH AND ACCOUNTS	AGI	DUMMY (AGI)	NET ASSETS
UP TO $2,000	0.777240	-0.052246	53.220	0.0166718
$2,000 TO $4,000	0.351396	0.019262	80.045	0.0074936
$4,000 TO $8,000	0.777724	-0.020367	102.881	0.0127374
OVER $8,000	0.120051	-0.013115	247.946	0.0181550

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (CASH)	INTERCEPT	R-SQUARE
UP TO $2,000	-206.93	152.834	97.0619	0.7437
$2,000 TO $4,000	-112.75	210.598	2.7298	0.2632
$4,000 TO $8,000	-570.49	399.976	18.3620	0.5695
OVER $8,000	-290.16	457.072	40.7170	0.3149

DEPENDENT VARIABLE-NET INVESTMENT VALUE

INCOME LEVEL	NET INVESTMENT VALUE	INTERCEPT	R-SQUARE
ALL LEVELS	0.779279	2040.73	0.1127

105

TABLE C-2

REGRESSION EQUATIONS USED FOR IMPUTATION OF BEST VALUES TO APPLICANT DATA BASE

Independent Students

DEPENDENT VARIABLE=EDUCATIONAL SOCIAL SECURITY

INCOME LEVEL	EDUCATIONAL SOC. SEC.	AGI	DUMMY (AGI)	NET ASSETS
UP TO $2,000	0.195394	0.74636	-372.3	-0.2052
$2,000 TO $4,000	0.177677	-0.63596	1548.6	-1.2111
OVER $4,000	-1.036327	1.41953	-8756.9	-1.1342

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (SS)	INTERCEPT	R-SQUARE
UP TO $2,000	-217.59	-1412.5	2348.20	0.1348
$2,000 TO $4,000	253.20	-63.1	470.78	0.2996
OVER $4,000	1363.47	-2870.8	4558.13	0.4238

DEPENDENT VARIABLE=EDUCATIONAL VA

INCOME LEVEL	EDUCATIONAL VA BENEFITS	AGI	DUMMY (AGI)	NET ASSETS
ALL LEVELS	0.0688744	0.0679462	1498.25	0.0709762

INCOME LEVEL	DUMMY (ASSETS)	DUMMY (VA)	INTERCEPT	R-SQUARE
ALL LEVELS	-459.02	622.329	479.612	0.2741

TABLE C-3

MEAN VALUES OF BEST, IMPUTED, AND REPORTED VALUES FROM THE STAGE III AND APPLICANT DATA BASES

Dependent Students

DATA ITEM—ADJUSTED GROSS INCOME

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$8,000 OR LESS	2103.1	2636.1	2636.0	1501.5	1996.2
$8,001 - $15,000	9177.1	9485.7	9486.3	8783.5	8964.0
$15,001 - $20,000	16178.6	16095.5	16095.5	15257.5	15063.2
OVER $20,000	23927.0	23983.4	23983.5	26599.8	26545.5

DATA ITEM—NON-TAXABLE INCOME

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$8,000 OR LESS	2121.73	2624.67	2627.49	1688.54	2303.33
$8,001 - $15,000	1965.75	2281.95	2281.95	2394.35	2752.61
$15,001 - $20,000	1114.50	1310.60	1312.51	2051.70	2298.31
OVER $20,000	870.38	1030.49	1030.50	1188.83	1311.73

DATA ITEM—NET HOME VALUE

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$8,000 OR LESS	10231.6	10788.9	10774.9	10362.4	11648.8
$8,001 - $15,000	12590.1	12765.5	12765.6	14968.9	15025.5
$15,001 - $20,000	14967.1	14702.9	14702.9	20402.6	19352.2
OVER $20,000	17491.1	18678.6	18678.6	28901.1	28631.6

DATA ITEM—TAXES PAID

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$8,000 OR LESS	82.20	48.97	49.51	67.53	71.23
$8,001 - $15,000	592.30	438.07	489.05	488.80	495.66
$15,001 - $20,000	1299.43	1300.99	1302.05	1308.29	1304.29
OVER $20,000	2481.85	2405.91	2407.61	3633.19	3538.28

TABLE C-3

MEAN VALUES OF BEST, IMPUTED, AND REPORTED VALUES FROM THE STAGE III AND APPLICANT DATA BASES

Dependent Students

DATA ITEM=DEPENDENT'S ASSETS

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$8,000 OR LESS	102.013	180.853	186.534	190.808	217.550
$8,001 - $15,000	83.886	137.228	137.220	171.098	179.970
$15,001 - $20,000	112.313	162.905	162.905	250.359	223.008
OVER $20,000	125.532	206.969	209.088	407.309	491.596

DATA ITEM=DEPENDENT'S INCOME

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$8,000 OR LESS	1130.01	1532.58	1532.59	980.00	1070.63
$8,001 - $15,000	1221.06	1559.56	1559.56	1208.58	1637.59
$15,001 - $20,000	1287.65	1612.78	1612.78	1027.72	1710.05
OVER $20,000	1286.19	1636.35	1636.32	1631.01	1873.71

DATA ITEM=MOTHER'S EARNED INCOM

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$8,000 OR LESS	2019.14	2008.31	2007.55	1379.32	1906.78
$8,001 - $15,000	5461.07	5855.89	5860.56	8883.69	5318.27
$15,001 - $20,000	7272.63	7422.80	7022.60	6142.72	6329.09
OVER $20,000	7026.57	6969.09	6962.60	6337.12	6059.24

DATA ITEM=NET INVESTMENT VALUE

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$15,000 OR LESS	823.90	1067.26	1071.11	1257.01	2087.93
$15,001 - $20,000	1172.95	1507.95	1498.80	2709.20	2056.77
OVER $20,000	1240.03	1512.20	1516.31	4131.61	4500.65

TABLE C-3

MEAN VALUES OF BEST, IMPUTED, AND REPORTED VALUES FROM THE STAGE III AND APPLICANT DATA BASES

Dependent Students

DATA ITEM=CASH AND ACCOUNTS

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$0,000 OR LESS	1011.40	865.43	868.20	820.76	898.89
$0,001 - $15,000	1091.98	1036.19	1045.36	1442.71	1493.85
$15,001 - $20,000	1366.24	1210.11	1211.10	2459.47	2330.17
OVER $20,000	1496.30	1408.46	1395.18	3116.19	2792.69

DATA ITEM=MEDICAL/DENTAL

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$0,000 OR LESS	222.891	244.016	243.921	227.007	246.522
$0,001 - $15,000	385.964	415.087	416.399	415.409	447.952
$15,001 - $20,000	556.418	574.637	574.086	581.744	589.036
OVER $20,000	[illegible].49	707.376	708.409	745.500	774.884

DATA ITEM=UNREIMBURSED TUITION

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
ALL LEVELS	85.4019	86.7288	86.7594	108.629	114.314

DATA ITEM=EDUCATIONAL SUP. SEC.

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$0,000 OR LESS	243.417	227.988	227.474	911.45	649.818
$0,001 - $15,000	237.470	172.692	172.713	1554.89	748.187
OVER $15,000	62.269	52.205	52.671	787.32	419.971

TABLE C-3

MEAN VALUES OF BEST, IMPUTED, AND REPORTED VALUES FROM THE STAGE III AND APPLICANT DATA BASES

Dependent Students

DATA ITEM=EDUCATIONAL VA BENEFI

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
ALL LEVELS	7.11389	16.0658	16.0511	25.6826	28.6828

DATA ITEM=NET BUSINESS/FARM VAL

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$15,000 OR LESS	3656.31	3927.82	3932.94	8174.94	8495.19
OVER $15,000	2103.72	1962.01	1979.44	3897.12	3759.94

DATA ITEM=FATHER'S EARNED INCOM

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$6,000 OR LESS	1196.4	1700.5	1706.2	1217.4	1967.6
$6,001 - $15,000	3663.1	4032.5	4031.0	3970.0	4300.3
$15,001 - $20,000	8763.4	9241.1	9259.9	8723.9	9144.5
OVER $20,000	16303.0	16784.5	16793.4	19668.5	20237.8

TABLE C-3

MEAN VALUES OF BEST, IMPUTED, AND REPORTED VALUES FROM THE STAGE III AND APPLICANT DATA BASES

Independent Students

DATA ITEM=ADJUSTED GROSS INCOME

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$2,000 OR LESS	410.0	729.3	761.4	278.2	781.7
$2,001 - $4,000	1726.3	1795.0	1793.0	1552.7	1660.0
$4,001 - $8,000	3892.3	3979.3	3975.3	4005.3	4113.5
OVER $8,000	10183.5	10087.8	10096.1	19389.3	19557.9

DATA ITEM=NON-TAXABLE INCOME

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$2,000 OR LESS	142.78	365.26	367.90	207.96	490.38
$2,001 - $4,000	1045.07	1216.12	1216.87	1301.72	1521.81
$4,001 - $8,000	1621.17	1804.07	1808.27	1819.93	1994.88
OVER $8,000	1172.69	1514.92	1513.38	1367.58	1555.87

DATA ITEM=NET HOME VALUE

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$4,000 OR LESS	390.86	440.07	449.24	484.46	524.40
$4,001 - $8,000	1232.96	1326.34	1327.90	1204.74	1339.41
OVER $8,000	3246.64	3326.81	3329.31	9088.88	9679.39

DATA ITEM=TAXES PAID

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$4,000 OR LESS	26.767	20.398	22.090	42.62	52.67
OVER $4,000	339.458	343.289	342.788	2132.32	2036.41

TABLE C-3

MEAN VALUES OF BEST, IMPUTED, AND REPORTED VALUES FROM THE STAGE III AND APPLICANT DATA BASES

Independent Students

DATA ITEM=SPOUSE'S EARNED INCOM

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
ALL LEVELS	1043.60	1136.75	1136.32	7999.32	8328.22

DATA ITEM=NET INVESTMENT VALUE

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
ALL LEVELS	68.7830	90.3650	90.8225	685.235	669.979

DATA ITEM=CASH AND ACCOUNTS

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$2,000 OR LESS	159.814	170.238	172.079	108.071	133.421
$2,001 - $4,000	151.715	158.905	158.858	127.101	133.638
$4,001 - $8,000	143.272	163.566	164.049	191.366	207.635
OVER $8,000	283.807	256.516	259.023	903.008	689.487

DATA ITEM=MEDICAL/DENTAL

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$8,000 OR LESS	79.135	85.053	85.228	61.645	70.099
OVER $8,000	202.553	218.808	219.104	368.080	386.397

DATA ITEM=UNREIMBURSED TUITION

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
ALL LEVELS	9.21284	9.69129	9.76517	67.4495	70.1938

BEST COPY AVAILABLE

TABLE C-3

MEAN VALUES OF BEST, IMPUTED, AND REPORTED VALUES FROM THE STAGE III AND APPLICANT DATA BASES

Independent Students

DATA ITEM=EDUCATIONAL SOC. SEC.

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$2,000 OR LESS	433.271	232.512	235.199	33.619	104.509
$2,001 - $4,000	188.063	97.977	98.194	229.105	121.144
OVER $4,000	109.721	57.443	58.011	334.607	540.490

DATA ITEM=EDUCATIONAL VA BENEFI

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
ALL LEVELS	133.219	114.832	114.709	280.775	327.113

DATA ITEM=STUDENT'S EARNED INCO

INCOME LEVEL	STAGE III REPORTED	STAGE III BEST	STAGE III IMPUTED	APPLICANT REPORTED	APPLICANT IMPUTED
$2,000 OR LESS	559.50	912.12	935.19	314.05	848.72
$2,001 - $4,000	1704.33	1932.39	1928.10	1578.15	1345.80
$4,001 - $8,000	3109.27	3414.21	3421.04	3417.73	3642.34
OVER $8,000	5756.81	6395.95	6391.55	7946.18	8673.66

APPENDIX D

TABLE D-1

COMPARISON OF NUMBER OF RECIPIENTS AND PROGRAM COSTS FOR THE 1982-83 PELL PROGRAM YEAR UNDER THE FULL AND FIVE DATA ELEMENT FORMULAE USING STANDARD REPORTED DATA WITHOUT TAXATION RATE ADJUSTMENTS

	Full Formula				Five Element Formula			
	ESTIMATED		COMPUTED EXPECTED AWARD		ESTIMATED		COMPUTED EXPECTED AWARD	
	NUMBER OF RECIPIENTS	% OF TOTAL RECIPIENTS	TOTAL COST	PERCENT OF TOTAL PROGRAM COST	NUMBER OF RECIPIENTS	% OF TOTAL RECIPIENTS	TOTAL COST	PERCENT OF TOTAL PROGRAM COST
TOTAL	2522705	100	2400965116	100	2717572	100	2620098393	100
TOTAL INCOME								
0 OR LESS	147501	6	161003469	7	148939	5	163084567	6
1-4,000	623400	25	720203903	29	628497	23	716200655	27
4,001-7,500	527031	21	564716739	23	539375	20	574980399	22
7,501-10,000	254096	10	277498975	11	266917	10	291144296	11
10,001-12,000	179806	7	184351905	7	193650	7	198835694	8
12,001-15,000	233699	9	215216344	9	255950	9	281770638	9
15,001-20,000	267650	11	212003444	9	324805	12	255214606	10
20,001-25,000	162665	6	97307244	4	205377	8	116093381	4
25,001-30,000	81508	3	37595049	2	100556	4	44019978	2
30,000 OR MORE	49348	2	17227164	1	53506	2	18802179	1

D-2

TABLE D-2

COMPARISON OF NUMBER OF RECIPIENTS AND PROGRAM COSTS FOR THE 1982-83 PELL PROGRAM YEAR UNDER THE FULL AND FIVE DATA ELEMENT FORMULAE USING ERROR FREE DATA WITHOUT TAXATION RATE ADJUSTMENTS

	Full Formula				Five Element Formula			
	ESTIMATED		COMPUTED EXPECTED AWARD		ESTIMATED		COMPUTED EXPECTED AWARD	
	NUMBER OF RECIPIENTS	% OF TOTAL RECIPIENTS	TOTAL COST	PERCENT OF TOTAL PROGRAM COST	NUMBER OF RECIPIENTS	% OF TOTAL RECIPIENTS	TOTAL COST	PERCENT OF TOTAL PROGRAM COST
TOTAL	2421948	100	2273058334	100	2595011	100	2422010034	100
TOTAL INCOME								
0 OR LESS	67331	3	72103607	3	68046	3	73489603	3
1-4,000	511554	21	568188474	25	515161	20	574838031	24
4,001-7,500	557096	23	574611825	25	571690	22	596175304	25
7,501-10,000	267539	11	277509303	12	278173	11	294705626	12
10,001-12,000	202377	8	200203886	9	214736	8	215505857	9
12,001-15,000	247066	10	216443173	10	267724	10	244690873	10
15,001-20,000	309228	13	234560120	10	348753	13	261047706	11
20,001-25,000	164676	7	90566541	4	203950	8	110545010	5
25,001-30,000	67109	3	28963111	1	86557	3	36605939	2
30,000 OR MORE	27973	1	9012294	0...	40221	2	14405906	1

BEST COPY AVAILABLE