The Student Follow-up Investigation of the 1979 Income Survey Development Program

Anthony M. Roman, Diane V. O'Brien, U.S. Bureau of the Census

I. Background

The Income Survey Development Program (ISDP) was a research and development program established in the mid-1970's by the Department of Health, Education and Welfare (HEW) in conjunction with the U.S. Census Bureau to prepare for the upcoming Survey of Income and Program Participation (SIPP). The SIPP is the new survey conducted by the Census Bureau designed to satisfy a wide variety of data needs concerning the economic situation of persons and families living in the United States. Data collection for the first SIPP survey, the 1984 Panel, began October 1983.

The major purposes of the ISDP were the same as the goals set out for the SIPP: to improve current estimates of income and income change; to extend the scope and precision of policy analyses for a wide range of Federal and State tax and social welfare programs, and to broadly assess the economic well-being of the population.[1]

The ISDP conducted four field tests. All were experimental in nature as different concepts, procedures, questionnaires and recall periods were tested. The 1979 Research Panel was the largest and most comprehensive research effort conducted by the ISDP.

The 1979 Panel was a nationwide household survey with a total sample of 11,800 households drawn from 130 Census primary sampling units (PSUs). Of this total, approximately 9300 cases were selected from an area sample and 2500 cases were drawn from list samples. Data collection began in February 1979 and ran through June 1980. One-third of the sample households were interviewed each month during the interview period. Information was obtained on household composition, labor force participation, various sources of money and nonmoney income, taxes, assets and liabilities, and other related topics.

The 1979 Panel included many controlled experiments which tested alternatives for basic survey design. The major tests conducted were: household versus individual questionnaire format; self versus proxy respondent rules; and 3-month versus 6-month respondent recall.

As part of the research effort to test respondent rules, one unresolved issue concerned proxy interviews taken for college students not living at their parents' address. In order to test the validity of information collected for this type of proxy interview, an experiment was conducted during the November and December interviews of the 1979 ISDP Panel. This experiment was called the Student Followup Investigation. This paper discusses the objectives, design, and field procedures used for the investigation, and some preliminary results of this experiment.

II. Purpose

Respondent rules during the 1979 Research Panel were to conduct a personal interview for each adult household member 16 years or older. If a self-response interview could not be obtained, the procedure was to accept a proxy interview from another household member who was knowledgeable about the absent person. In this survey, as in other Census surveys, students were considered as members of their parents' households until they established a permanent residence elsewhere. Therefore, the usual procedure for students living away from home while attending school was to treat them as household members who were temporarily absent and obtain proxy interviews from other members of their parents' household.

The fourth interview questionnaire (Wave 4) used during the 1979 Panel contained a special set of questions concerning post secondary educational enrollment and expenses. Thus, this interview seemed especially appropriate for studying the quality of proxy interviews for students, as compared to the student's self interview.

In order to measure the accuracy of information taken from proxy interviews for students living away from home, the fourth interview was first obtained by proxy at the parents' household, and then by self interview at the student's school residence. This self-response interview is referred to as the student followup interview.

There were two basic purposes for conducting the Student Followup Investigation:

1) To obtain the most complete and accurate information possible for items in the Education Expenses section of the Wave 4 questionnaire (such as school enrollment, tuition, fees, and living expenses), and

2) To determine whether proxy respondents at the sampled address are able to provide reliable information on labor force participation, income, education expenses and enrollment for students living away from home. This experiment would be conducted by comparing the information obtained from the proxy interview taken at the parents' home to the self-response interview taken at the student's school address.

III. Procedures

The fourth interview questionnaire used in November and December was designed to identify students living away from their usual residence while attending school. Only students who were actually staying at their school residence (either a dormitory, fraternity house, apartment, etc.) during the time of the November or December interview were eligible for followup.

Census interviewers were instructed to call their regional office within 24 hours of a household interview in which a proxy interview had been administered for a student who was absent and living away at school. The interviewer would then provide the office with the student's name and school address.

Census regional offices were responsible for the control and assignment of the student followup cases. The rules for assigning the cases were essentially the same as the ISDP rules for movers. If the student's school address was within 50 miles of an ISDP PSU, the office assigned the case for interview. As soon

as possible thereafter, an interviewer visited the student for an interview. Regional offices were instructed to always employ a different interviewer for the student's interview in order to eliminate any interviewer bias. Additionally, interviewers were instructed to accept only self-response interviews at the student's school address; no proxy responses from roommates or friends were allowed.

IV. Field Results

The analytic universe for the study was the totality of students in the 4th Wave of the 1979 Panel who usually lived away from home and were attending post secondary schools.[2] There were 443 such students identified. Of these, 117 (26.4 percent) were not eligible for interview since the school residence was more than 50 miles from an ISDP PSU and 54 (12.2 percent) were not eligible because the student was staying at home during the time of the 4th Wave interview.

Of the 272 cases assigned, 202 student followup interviews were obtained yielding a response rate of 74.3 percent. Of the 70 noninterviews, 6 were cases in which the parents refused permission for the interviewer to contact the student.

The major reason for the noninterviews was that many students were not staying at their school address (because of Thanksgiving, Christmas and semester breaks) by the time the interviewer received the followup assignment. Although interviewers were allowed until the first week of December to obtain the followup interviews for students identified in November and until the second week of January for students identified in December, many students remained on some type of break later into December and January. This proved to be an inappropriate time of year for conducting interviews with students at their school address. However, in the case of the 1979 Panel, we overlooked this factor in the survey design in order to conduct the experiment in conjunction with the Education Expenses questions, which were set beforehand for the Wave 4 interview.

A recommendation for future studies involving students interviewed at their school address is to obtain the school address in a previous wave's interview. This would allow interviewers more time to contact the student.

V. Preliminary Findings

A. Data Set Creation

The first task in analyzing these data was the creation of a data set of matched responses from the student followup questionnaire and the proxy questionnaire administered during Wave 4 of the ISDP. During the matching process, 35 students (17.3 percent) could not be matched to the Wave 4 ISDP File. Attempts to reconcile the mismatches were unsuccessful. In all but one instance, the most basic identifiers for these 35 students did not exist on the Wave 4 ISDP File. Due to the time elapsed from the initiation of this followup study to the creation of the analysis data set, it has been extremely difficult to find out why these

mismatches have occurred. Future studies should be aware of these problems and prepare for them. Omitting the 35 mismatches resulted in a data set of 167 matched responses. These data are analyzed in this report. In all but two instances, the variables analyzed are direct responses to questions on the ISDP form (i.e., they are not in any way computed). The only exceptions are "usual hours worked per week at all jobs" and "total pay before deductions from all jobs last month". These two variables are computed by summing the response from each reported job.

B. Relationship of Student to Proxy Respondent

The relationship of the student to the respondent serving as his/her proxy can be determined in most cases through their relationships to the household reference person. The reference person is that household member who is stated as owning or renting the residence. Table 1 indicates that in 84.4 percent of the cases, the proxy was a parent of the student. This follows the expected pattern.

C. Wage and Salary Comparisons

The ISDP questionnaire was divided into several sections. One section was designed to identify receipt of income types while other sections obtained amounts. Persons were asked a series of wage and salary questions if they indicated in the recipiency section of the questionnaire that they worked at a job or business. One wage and salary record was created containing responses to the set of wage and salary questions for each job named. Thus, if a student had only one employer, a wage and salary record should have been created with the student's responses while another wage and salary record should have been created with the proxy's responses. The reference period used in the ISDP was the previous 3 months, but the wage and salary records were created on a job basis. Therefore, a reported job could have been held at any time during the 3-month reference period. In examining the 167 matched cases of self and proxy responses, the following breakdown of wages and salaries was observed:

    83 had at least one self and one proxy record
    53 had neither a self nor a proxy record
    27 had a self but no proxy record
    4 had a proxy but no self record

If one assumes that the self response is correct, then the proxy failed to identify a job held by the student in 27 cases (24.5 percent). This appears to be rather substantial and indicates a potential source of underreporting of wages and salaries with proxy response. The 4 cases in which a proxy record exists while no self record exists may be interpreted as a potential source of misreporting wages and salaries under proxy response.

In attempting to analyze particular wage and salary questions of interest, several conditions must be kept in mind. While 83 matched cases exist with both a self and proxy wage and salary record, the number of cases available for making comparisons for any particular question may be less. There are two primary reasons for this: 1) one interview may have proceeded in a fashion

which asked the question of interest and a response was coded, while the other interview proceeded in a fashion which did not ask the question (i.e., due to the various possible skip patterns within the questionnaire), and 2) even though the question of interest may have been asked during both interviews, one may have resulted in a valued response[3] while the other did not. Valued responses are important in evaluating the quality of data obtained in a survey. They indicate both knowledge by the respondent of the investigated subject matter and willingness to cooperate in the survey.

With this in mind, the percentages of coded responses which were valued (i.e., given that a question was asked, the number of times it resulted in a valued response) are presented in Table 2. It is seen that for several wage and salary questions, it is more likely that a self respondent will give a valued response. This is particularly evident with the "usual hours worked per week" and "hourly rate of pay" variables. In all but one instance, when a valued response was not given, a "don't know" was the recorded response.

Table 2 also presents the mean value of self responses for seven wage and salary variables for three particular categories:

1) the proxy could not identify that the student had a job (i.e., no proxy wage and salary record existed but a self wage and salary record did exist)

2) the self response was valued while the proxy response was not (e.g., the proxy most likely responded "don't know"), and

3) both self and proxy responses were valued.

This table demonstrates that a pattern appears to exist in which proxies best identify jobs at which students earn the most money or work the most hours. The smaller the earnings or hours worked, the more likely the proxy will either not be able to identify the job or not be able to answer detailed questions about the job.

Several points should be noted concerning Table 2. The usual hours worked per week may seem rather high for student jobs. This is due to the reference period for these questions extending back into the summer months. Therefore, summer jobs in which the student may have worked 40 or more hours per week will be included in these summaries. This also explains the decreases in total monthly pay from three months ago to last month. Also, it is impossible to compute total monthly pay by using the usual hours worked per week and regular hourly rate of pay. This is because the values presented in these tables are means and concern the student's primary job. One student's primary job may have been three months ago while another's may have been last month.

The final table, Table 3, presents comparisons of the self and proxy valued responses. It should be noted that the estimated variances used in computing these confidence intervals do not take into account any sample design effects. The reason is that this analysis is considered preliminary and will be used to decide if a more lengthy detailed

analysis seems warranted. The net result is that the intervals in Table 3 should be considered conservative, while conclusions of significant differences should be considered liberal. Computation of design effects may add a small degree of accuracy to results from this study, but it should be noted for future studies, that increased emphasis on obtaining responses from all sample students and their proxies would greatly enhance the accuracy of results. Of the seven wage and salary variables analyzed, two showed a significant difference at the .05 level. These were "usual hours worked per week" and "regular hourly rate of pay", both for the student's primary job. In both instances, the proxy gave the larger valued mean response. It is interesting to note that for "usual hours worked per week at all jobs", the mean self and proxy responses are not significantly different. This raises the question of the proxy and student possibly identifying different jobs as being primary.

D. Education Expenditure Comparisons

All 167 matched cases had both a self and a proxy education expenditures record, but 61 of these records were unavailable for this preliminary analysis. This was due to a flaw discovered in the manner in which the Wave 4 ISDP data were processed. Only rekeying of the questionnaires could retreive the data and this was deemed unwarranted at the present time. Therefore, 106 matched records were available for analyzing education expenditures.

Table 2 again presents the percentage of responses which were valued. It is obvious that a valued response is much more likely from a self respondent than a proxy respondent. This seems understandable for all variables except "amount paid by family on tuition and fees" since the other variables involve expenditures most likely handled directly by the student. In every instance that a valued response was not given, "don't know" was the recorded response.

Table 2 displays the mean value of self responses both when the proxy has a valued response and also when the proxy response is "don't know". Three of the four variables considered do not appear to differ substantially between these two categories. Only the "amount paid by family on tuition and fees" exhibits a rather large difference with the mean self response being greater if the proxy has a valued response. This is consistent with the wage and salary results in that the more expensive the tuition, the more the proxy is likely to know about the amount. It may also help explain why so many "don't knows" were given by proxies in response to this question. Perhaps when the amount of tuition is low, the student is more likely to be directly involved in its payment(e.g., the student may pay the tuition from support supplied by the parent).

Table 3 again presents results of comparisons of self and proxy valued responses. Two of the four variables showed a significant difference at the .05 level. They were "academic credit hours taken this term" and "cost of course materials". In both instances, the mean proxy

response was larger. The mean proxy response was also larger for "amount paid on tuition and fees" but with a large estimated variance, a statistically significant difference could not be detected.

E. Other comparisons

Two additional areas were investigated in this preliminary analysis. The first was educational assistance. Of eight assistance categories, only two had enough reported cases to analyze questions concerning amounts received. These were: Basic Educational Opportunity Grants (31 cases) and Government Scholarships, Fellowships, Etc.(11 cases). The results of comparisons of self and proxy valued responses are shown in Table 3. No significant differences in mean amount received were found for any of the assistance variables.

The last area investigated was receipt of interest income. Reporting of interest was handled in the ISDP questionnaire in the same manner as wages and salaries. That is, a person was asked a series of questions regarding amounts of interest if they indicated receipt of interest income in the recipiency portion of the questionnaire. For the 167 matched cases, the reported interest was as follows:

   104 cases had both a self and proxy report
    30 cases had neither a self nor proxy report
    27 cases had a self but no proxy report
     6 cases had a proxy but no self report

Assuming the self response is correct, the proxy failed to identify that the student would have interest income in 27 cases (20.6 percent). Although this appears to be a large problem, interest income is poorly reported for all people. For example, in the 104 cases in which both the self and proxy respondent reported receipt of interest earned on the student's own accounts, 61.0 percent of the coded self responses were "don't know" while 81.5 percent of the coded proxy responses were "don't know". Considering the question on interest earned on the student's shared accounts, 69.4 percent of the coded self responses and 80.0 percent of the coded proxy responses were "don't know". Obviously, it appears that the quality of interest data for students is suspect regardless of whether a self or proxy interview is conducted.

VI. Conclusions

The aim of this preliminary analysis was to examine the self and proxy student data in order to decide if a more extensive investigation (e.g., effects of accepting proxy responses on overall survey estimates) seemed warranted. Any inferences drawn from these data should keep in mind that the estimated variances did not reflect any sample design effects and that the size of the data set is quite small. Indeed, most comparisons were based on less than 100 observations. Still, this study is unique and although somewhat flawed in administration and implementation, it is possible to make certain general remarks. When valued responses are available from both the self and proxy interviews, the quality of the proxy responses appears to be generally quite good.

Substantially more data would be needed to derive better estimates of the difference between self and proxy response and to narrow the confidence intervals around these estimates.

A problem that does appear to exist is in obtaining a valued proxy response. Quite often, a proxy cannot identify a particular source of student income (e.g., wages and salaries) and even if they can identify it, they are more likely to respond "don't know" to the particulars about that source. A trend does seem to exist that the larger the income or expense, the better the proxy response becomes. Still, this implies that by using proxy responses, the lower range of income or expense amounts are more likely missed.

Finally, the main issues involved in interviewing students away from home are the impact of accepting proxies on overall survey estimates and the differential costs involved in obtaining self responses. Since no cost data are available from this study, an estimate of the additional amount required in obtaining self responses cannot be computed. It may be possible to make some very general comments about the potential impact of accepting proxies on overall survey estimates. Students living away from home make up less than 3 percent of the overall ISDP sample. With this in mind and the fact that results from this study indicate that proxies are more likely to miss only the smaller expense and income amounts, it may appear unlikely that overall survey estimates will be strongly affected. Still, the limitations of the sample involved in this study must be considered in any statement of results. For instance, students living more than 50 miles from an ISDP PSU were omitted from consideration. Also, problems were encountered in matching students to proxies and in losing some survey data due to a processing flaw. The effect that these students could have had on results from this study is unknown. In concluding, further detailed investigation of this particular data set is not recommended due to the limitations in the size and composition of the sample. Future study may lead to stronger results but based upon this preliminary investigation, it is recommended that while the self-proxy student issue should not be forgotten, it should not occupy a high place on the SIPP research agenda.

FOOTNOTES
1 Research Triangle Institute. 1983. The 1979 ISDP Research Panle Documentation. National Technical Information Service, Washington, D.C.
2 Since Wave 4 of the 1979 Panel was administered over a two month period, only two-thirds of the 11,800 household sample was interviewed, making the Wave 4 sample size approximately 8,100 households.
3 Throughout this report, the term valued response is used to imply any response with a legitimate value for the question asked. Valued responses do not include refusals, don't knows, or responses whose value is considered out of range or in some other manner erroneous.

Proxy:

student:	ref. person	spouse of ref. pers.	child of ref. pers.	other rel. of ref. per.	unknown
ref. pers.	–	0	0	2	1
child of ref. pers.	60	81	6	0	11
other rel. of ref. pers.	3	1	2	0	0
Total	63	82	8	2	12

Table 2: Results for Wage & Salary and Education Expenditure Variables

Variable	% of coded cases which were valued:[1]		Mean value of self responses when:[2]		
	self response	proxy response	proxy could not identify that student had a job	proxy did not give a valued response	both self and proxy had valued responses
Wage and Salary					
Usual hours worked per week at primary job	98.7% (n=76)	76.4% (n=55)	22.30 hrs (n=20)	21.11 hrs (n=38)	35.60 hrs (n=37)
regular hourly rate of pay at primary job	100.0 (n=66)	73.3 (n=75)	$3.17/hr (n=16)	$3.46/hr (n=18)	$3.39/hr (n=48)
total pay before deductions from primary job 3 months ago	100.0 (n=60)	100.0 (n=51)	$111.81 (n=16)	$246.48 (n=27)	$378.42 (n=33)
total pay before deductions from primary job 2 months ago	100.0 (n=61)	100.0 (n=50)	$32.19 (n=16)	$97.82 (n=28)	$138.33 (n=33)
total pay before deductions from primary job last month	100.0 (n=61)	98.0 (n=50)	$37.25 (n=16)	$74.83 (n=29)	$100.00 (n=32)
usual hours worked per week at all jobs	–	–	25.76 hrs (n=21)	24.75 hrs (n=37)	41.37 hrs (n=38)
total pay before deductions from all jobs last month	–	–	$94.47 (n=17)	$131.10 (n=63)	$114.53 (n=51)
Education Expenditures					
academic credit hours taken this term	98.0 (n=101)	50.0 (n=102)	–	14.88 hrs (n=49)	14.80 hrs (n=50)
amount paid by family on tuition and fees this term	95.0 (n=100)	74.5 (n=102)	–	$469.25 (n=24)	$1004.10 (n=71)
cost of course materials this term	100.0 (n=100)	46.1 (n=102)	–	$105.59 (n=54)	$98.75 (n=44)
amount of monthly rent and utilities	100.0 (n=30)	66.7 (n=33)	–	$148.67 (n=9)	$131.14 (n=21)

1   In these two columns, the numbers in parentheses are amounts of coded cases.
2   In these three columns, the numbers in parentheses are amounts of valued responses.

Variable	Mean Self Response	Mean Proxy Response	Difference	95% Confidence Limits [1] Lower	Upper
Wage and Salary					
usual hours worked per week at primary job	35.60 hrs	40.57 hrs	−4.97 hrs*	−9.03 hrs	−0.91 hrs
regular hourly rate of pay at primary job	$3.39/hr	$3.54/hr	−$.15/hr.*	−$.29/hr	−$.01/hr
total pay before deductions from primary job 3 months ago	$378.42	$336.09	$42.33	−$56.80	$141.46
total pay before deductions from primary job 2 months ago	$138.33	$121.52	$16.81	−$30.79	$64.41
total pay before deductions from primary job last month	$100.00	$106.56	−$ 6.56	−$22.56	$ 9.44
usual hours worked per week at all jobs	41.37 hrs	40.55 hrs	0.82 hrs	−3.55 hrs	5.19 hrs
total pay before deductions from all jobs last month	$114.53	$114.38	$.15	−$21.65	$21.35
Education Expenditures					
academic credit hours taken this term	14.80 hrs	16.00 hrs	−1.20 hrs*	−1.93 hrs	−0.47 hrs
amount paid by family on tuition and fees this term	$1004.10	$1157.63	−$153.53	−$591.87	$284.81
cost of course materials this term	$98.75	$120.84	−$22.09*	−$39.01	−$5.17
amount of monthly rent and utilities	$131.14	$121.38	$9.76	−$16.87	$36.39
Education Assistance					
BEOG assistance received 3 months ago	$299.61	$381.74	−$82.13	−$281.99	$117.33
BEOG assistance received 2 months ago	$194.48	$301.24	−106.76	−$302.28	$ 88.76
BEOG assistance received last month	$ 50.57	$ 14.71	$35.86	−$ 10.11	$ 81.83
Government scholarship assistance received 3 months ago	$ 87.55	$107.73	−$20.18	−$ 63.27	$ 22.91
Government scholarship assistance received 2 months ago	$154.75	$219.75	−$65.00	−$151.52	$ 21.52
Government scholarship assistance received last month	$ 8.00	$ 28.50	−$20.50	−$ 66.87	$ 25.87

* implies difference is significant at the .05 level

1 These limits are based on variance estimates which do not take sample design effects into account.