

THE QUALITY OF PROXY REPORTS ON THE CURRENT POPULATION SURVEY (CPS)¹

Brian A. Kojetin, Bureau of Labor Statistics and Paul Mullin, Inst. of Gov. and Public Policy, Univ. of Illinois
Brian A. Kojetin, BLS, 2 Massachusetts Avenue NE, Washington, DC 20212

Keywords: Proxy Reporting, CPS

What do people living in the same household know about each other? Do they, for example, know the employment status, work and vacation hours, and wages of other household members? Although household surveys frequently ask such information about each member in the household, it is a daunting and expensive task to obtain all of this information firsthand from each person. Therefore, many household surveys accept proxy reports from one person for all household members. In the case of the Current Population Survey (CPS), proxy reporters typically provide about 50% of the labor force data. This level of reliance on information obtained by proxy may well lead one to be concerned about the quality of the data that proxies supply.

In a review of the literature on proxy reporting Moore (1988) found little evidence that self reports were inherently superior to data provided by proxy. Rather than being a ringing endorsement of proxy reporting, Moore's review articulated the methodological weaknesses of much of the relevant literature on proxy reports, and the paucity of reliable information by which to accurately assess the quality of self and proxy reports.

More recently researchers have been approaching the study of self and proxy differences from theoretical perspectives drawn from cognitive and social psychology to provide a framework for understanding the differences in self and proxy reports. Furthermore, several well-controlled experimental studies have been conducted to examine processes underlying differences in self and proxy reports. For example, Schwarz and Wellens (1994) draw on social psychological theory and research on actor-observer differences to derive a set of hypotheses for differences in self and proxy reports. In a series of experimental studies they provide evidence that proxy reporters rely more on general dispositional information when making judgments about another person. Seymour Sudman and his colleagues (Menon, Bickart, Sudman & Blair, 1995; Sudman, Bickart, Blair, & Menon, 1994) have also conducted a program of research studies aimed at understanding the cognitive processes that underlie self and proxy reports and that help explain the convergence and divergence of

self and proxy reports. In a series of studies Sudman et al. (1994) demonstrated that higher participation and more discussion were reliably related to the convergence of self and proxy reports. In a recent set of studies (Menon et al., 1995), they also showed that the more participation and discussion that occurs between self and proxies, the more the cognitive strategies used by each is similar and the greater the convergence of their reports.

In the present study we sought to extend the research on proxy reporting to the specific context of a large, well-known government survey, the Current Population Survey (CPS), and to examine additional factors in understanding the convergence of self and proxy reports that are apparent in this context and also may generalize to other survey contexts as well. We have focused on the ways in which the proxy is likely to learn about the labor force activities of the target person and factors underlying this knowledge acquisition process. One would expect, and the above mentioned research has shown, that some ways of knowing, such as participation and discussion, should lead to greater convergence between self and proxy reports. We believe it is also important to note that the underlying context and perhaps even the motivation for a proxy to acquire knowledge about another person is the relationship that exists between the two people. In most cases, people living in the same household have some kind of relationship with each other, and often these relationships are the closest ones that the person has. There is also considerable variability in these relationships, but the more time household members spend together and the more activities they engage in together, over a longer period of time, the more knowledge they are likely to have about one another. Furthermore, specialization frequently occurs within relationships such that certain household members may know more about other household members regarding some specific domains. For example, one household member may take more responsibility for paying the household bills, and this person is likely to place greater importance on knowing the activities of other household members that are relevant to that task. Therefore, one would expect that proxy reporting would be better for

¹ The opinions expressed in this paper are those of the authors and do not necessarily reflect the opinions or policies of the Bureau of Labor Statistics. The authors wish to thank Oak Ridge National Laboratory for their programming of the instrument and the Institute of Social Research at the University of Tennessee for data collection.

domains that are particularly relevant or important to the proxy.

There are additional aspects of this research that differ from much previous research on proxy reporting. For example, unlike many researchers studying the quality of data provided by proxy reporters for a wide variety of questions including frequency and attitudinal measures, we are not only interested in relative agreement, i.e., do self and proxy reports correlate well, but are also concerned with the absolute values that each provides. When we ask a respondent about the number of hours a household member worked or wages earned, the actual value has meaning, and two reports that consistently differ is problematic, regardless of the high correlation between the reports. Researchers have also frequently utilized only married couples or even college roommates in studies of self and proxy reporting. Because any responsible adult may serve as a respondent in the CPS, we are also concerned about the proxy reports of parents for children and vice versa, as well as other relatives and non-relatives within the household.

The purpose of the present study was to first examine the agreement between self and proxy reports involving important labor force activities using survey questions from the Current Population Survey. Our second goal was to examine factors that we expected would be related to self-proxy agreement. Specifically, we included measures of the interaction, communication, and relationship between self and proxy reporters. We also sought to extend previous research by looking at other modes of communication in addition to discussion and participation and to examine the proxy's use of general knowledge of another household member.

While some may argue that it might be preferable to have external data to compare accuracy of self and proxy reports, we have employed self and proxy agreement as our criteria of the quality of a proxy report. Although self-reporters are fallible human beings who may forget or may intentionally not provide complete information about their own labor force activities, they are often the best criteria that can be reasonably obtained, and they also represent the only real alternative to proxy reports for survey researchers conducting household surveys.

Method

Overview of Design

All eligible members from ninety-seven households completed a computerized self-administered questionnaire that included questions from the CPS. Household members completed interviews independently and simultaneously on individual personal computers, reporting information for themselves and every other eligible member of their

household. Only household members who were at least 16 years old, who would have been acceptable as a respondent for the CPS, participated in the study. For each proxy report about another household member, respondents also answered questions concerning how they learned about that person's labor force activities and their recent interaction with and knowledge of that person's labor force activities.

Subjects

Data collection was done by the Institute for Social Research at the University of Tennessee. Initially, they conducted a random digit dialing telephone survey of 400 Knox county residents. Questions were asked to determine the size of the household and the willingness of members to participate in the study as well as demographic questions concerning the educational levels and race of the household members. A sample of households was drawn from this initial selecting households of different sizes and characteristics. Because we sought a relatively equal number of two, three and four person households and there were insufficient numbers of three and four person households from the RDD survey, extensive efforts were made by the researchers at the University of Tennessee to recruit families with the desired characteristics. Some demographic characteristics of the sample is shown in Table 1. Data from three households were lost completely due to software or hardware problems. Although attempts were made to obtain a sample that reflected the diversity and characteristics of Knox County, Tennessee, this sample is not representative, and was not intended to be representative.

The Current Population Survey

The CPS is the primary household survey for determining labor force status in the United States. In the present study a computerized self-administered instrument containing all of the major CPS questions as well as additional measures was programmed by Oak Ridge National Laboratory.

The CPS contains a variety of questions requiring different kinds of responses. Many questions impose a strict, recent specified time period, viz. the last week, for consideration and should elicit recall of more specific instances. For example, "LAST WEEK did do any work for pay?" Other questions draw upon more general knowledge and inferences about usual behavior, such as "How many hours a week does usually work?" Because of the skip patterns, only proxies and self-reporters who went down the same path can be compared for many questions. For the purposes of the present paper, we will focus on CPS questions on whether the person did any work for pay in the last week, their usual hours, their actual hours, their usual

earnings, and whether the person looked for work in the last week (if they did not have a job).

Measures of Communication

We measured the source of information for major CPS questions by asking proxy's how they learned about the labor force activities or information of another household. For each major CPS question, proxies were presented with a list of possible ways they could have learned about the target person's activity or status, and they were to select as many of them as applied. The possibilities included: participation with the target, conversation with the target, first hand observations, hearing from others, reading information, and general knowledge of the target.

Measures of Relationships

In this study we included modified measures that were designed to measure important dimensions of interpersonal relationships (Berscheid, Snyder, & Omoto, 1989). Household members indicated estimated the amount of time they spent with each other during the past week and whether they engaged in a variety of different activities with each other during the past week. We also collected information about the type of relationship, for example, spouse, parent-child, and child-parent, etc.

Results

Overview of Data Analyses

The present paper represents an initial report from a rich data set. The preliminary results of our research are presented in three sections. In the first section, we examined self and proxy agreement for several important questions from the CPS. The next section includes an examination of how well the proxy's mode of learning about the target person is related to self-proxy agreement. In the last section, we examine how the proxy's relationship with the target person and their self-rated knowledge about the target person's labor force activities are related to self-proxy agreement.

Levels of Self and Proxy Agreement

There were a total of 582 proxy reports to compare with the self-reports (slightly less on some items with missing data). A total of 574 matched self-proxy dyads reported sufficient information to obtain labor force classification and the results are shown below. Overall, 84.0% of the classifications were exactly the same and 87.5% of the classifications were the same for in-the-labor-force versus not-in-the-labor-force. The unemployment rate reported by all self-reports was 11.9% and the unemployment rate using all proxy reports was 12.8%.

Labor force classification is a summary classification obtained from answering several CPS questions. We will focus on the agreement for some of those questions as well as questions asked of employed persons. One of the first and most important questions

asked in the CPS is whether the person did any work for pay during the last week. Overall, 83.1% of the self and proxy responses were the same; however, because answers to this item initiated different branching, only the 339 proxy-self pairs who answered this question the same followed the same path from here onwards.

Respondents who did work for pay are routed to a series of questions about the hours they worked, their occupation, industry, and earnings. Specifically, in the redesigned CPS respondents are first asked how many hours they usually work and then are asked questions about hours they took off last week or overtime hours they worked last week and finally how many hours they actually worked last week. The self and proxy reports for usual and actual hours worked are shown in Table 2. Proxies reported that the target person usually worked significantly fewer hours than the self reported. However, this difference does not appear for actual hours worked last week. The correlation between self and proxy reports is around $r = .70$, $p < .01$, for both usual and actual hours. In addition, there was 85.6% agreement on the classification of usual part-time or full-time status, and approximately 75% of the self-proxy dyads also gave the same answers to the questions on whether they took off any hours last week or worked any overtime hours.

A series of items on the CPS also inquire about earnings. All respondents are given the choice of what time period is most convenient for them to report earnings: hourly, weekly, biweekly, monthly, or annually. Thus, a person giving a self report may give the amount from their biweekly paycheck and a proxy reporter may give that person's annual salary. To compare the self and proxy reports of earnings all time periods were converted to weekly earnings. There were no significant differences in the level of earnings reported by self and proxy reporters as can be seen in Table 2, and the correlation between self and proxy reports was $r = .80$, $p < .01$.

Respondents who did not do any work for pay the previous week and did not report having a job are asked if they have done anything to find work in the past four weeks. This question, in combination with others, is used to determine whether the person is unemployed or not in the labor force. Overall, 69.9% of the self-proxy dyads answering this question gave the same answer

Communication Mode and Self-Proxy Agreement

In the next phase of analyses we sought to examine factors that might account for differential agreement of some self-proxy pairs relative to others. In this section we focus on the communication mode the proxy had for learning about particular labor force activities of the target person. As can be seen in Table 3, some ways that proxies said that they learned about the labor force activities of the target person were significantly related

to agreement with the target person. Specifically, proxies who reported that they noticed the target person's activities or knew this to be generally true of them were more likely to agree with the target persons report that proxies not reporting these communication modes.

We next examined whether communication mode was related to the relative agreement of self and proxy reporters for hours and earnings. Table 4 shows the correlations of the self and proxy reports for proxies reporting different communication modes. Findings for actual and usual hours were very similar, so only results for actual hours are reported for the rest of this paper. Proxies who reported that the target person told them this information, who read something, or who noticed the target person's activities agreed more with the target person's report for hours or earnings than did proxies who did not report learning through these communication modes.

Relationship Characteristics and Knowledge and Self-Proxy Agreement

We further examined the degree to which aspects of the relationship between the self and proxy were related to their agreement about labor force activities. We chose spouses as a comparison for the other relationship types because they are common proxy reporters and are expected to be quite accurate. As can be seen in Table 5, there were differences in self-proxy agreement for different relationship types. Specifically, proxies who were not relatives or related in some other way than parent, child or sibling agreed less with self-reports on whether the target person worked for pay last week than proxies who were spouses. Proxies who were siblings or relatives other than parents or children agreed less with the self-reports of the target person on whether he or she did anything to look for work in the last four weeks.

We next examined whether the amount of time and the number of activities the self and proxy were together for and the proxy's self-rated knowledge of the labor force activities of the target would be related to self-proxy agreement. Although we performed statistical tests using these relationship and knowledge variables as continuous, for purposes of illustration here we performed median splits on each of these variables to distinguish low and high levels of interaction and knowledge. As can be seen in Table 6, proxies who said they knew more about the work-related activities of the target person actually agreed more with the target person's self-report on whether they did anything to look for work in the past four weeks.

To examine whether relationship type was related to self-proxy agreement for hours and earnings, correlations between self and proxy reporters were computed separately for each relationship type and are

shown in Table 7. It can be seen in Table 7 that children acting as proxy reporters agreed less with the target person's self-reports than spouses for hours and earnings, and siblings serving as proxies agreed less with the target's self-reports for hours than spouses did.

Similarly, the correlations of self and proxy reporters for hours and earnings are shown separately for proxies with high and low levels of interaction and knowledge in Table 8. Proxies who did more activities together with the target person showed higher correlations with the target person's self-reports than proxies who engaged in fewer activities. Proxies who said they know more about the target person's work related activities showed a higher correlation with the self-reports of earnings than proxies who said they knew less about the target person's work-related activities.

Conclusions

The present paper contains some initial findings from a large study on the quality of proxy reports on BLS surveys. We sought to examine both the levels of self and proxy agreement for a variety of important questions from the CPS as well as some possible factors that would predict which proxy reporters would agree most with the self-report of the target person. Although we found relatively high levels of self-proxy agreement and correlations compared to many studies of proxy reporting, there is still sufficient room for concerns about data quality because of the use of these data and the prevalence of proxy reporting in the CPS. In examining communication modes, we found that discussion and participation were not as important in this context as some other modes of communication or reliance on general knowledge. This may be due in part to other studies not examining some of the modes included here as well as the nature of the questions being asked. We expected that relational factors would be more directly related to self-proxy agreement than was demonstrated here. It may be that these factors are more indirectly related through the use of communication mode and general knowledge.

Limitations

The structure of skip patterns in the CPS limits some of the comparisons that we are able to make between self and proxy reporters. Proxies who answer the work for pay question differently than the target person can not be compared on any other questions because they go down different paths and some paths have too few respondents to make meaningful comparisons. Furthermore, any comparisons for hours and earnings implied some prior agreement between the self and proxy and may be restricting the range of some of our predictors of self-proxy agreement. We do not have data from this study reconciling differences between self and proxy reports which might help us arrive at an accuracy measure. We also did not obtain external data

such as paystubs or information from the employer. The present study utilized a computerized self-administered version of the CPS, which is an interviewer conducted survey in the Field. Errors may enter into this self-administered version that would not be present in an the actual CPS interview.

Future Directions

All of the participants in this study also completed self and proxy reports for some expenditure categories from the Consumer Expenditure Surveys sponsored by BLS. We plan to examine these same factors in those data and compare their importance in the context of those questions about different purchases that were made, which are highly specific and have differing time boundaries.

References

Berscheid, E., Snyder, M., and Omoto, A. M. (1989). The relationship closeness inventory: Assessing the closeness of interpersonal relationships. *Journal of Personality and Social Psychology*, 57, 792-807.

Menon, G., Bickart, B., Sudman, S., & Blair, J. (1995). How well do you know your partner? Strategies for formulating proxy-reports and their effects on the convergence to self-reports. *Journal of Marketing Research*, 32, 75-84.

Moore, J. C. (1988). Self/proxy response status and survey response quality: A review of the literature. *Journal of Official Statistics*, 4, 155-172.

Schwarz, N. & Wellens, T. (1994). *Cognitive dynamics of proxy responding: The diverging perspectives of actors and observers*. Unpublished manuscript, U. S. Census Bureau, Washington, D.C.

Sudman, S., Bickart, B., Blair, J., & Menon, G. (1994). The effect of participation level on reports of behavior and attitudes by proxy reporters. In N. Schwarz and S. Sudman (Eds.) *Autobiographical memory and the validity of retrospective reports*. New York: Springer-Verlag.

Table 1. Description of the Sample

Household Characteristics N = 97
39 with 2 respondents; 30 with 3 respondents; 28 with 4 respondents
57% Household Income from 20,001 to 60,000
Respondent Characteristics N= 280 respondents
Ages ranged from 16 to 85; mean 33.4 years; median 29 years.
54% Females
89% White
60% High School Graduate or less

Table 2. Self and Proxy Agreement on Hours Worked and Earnings.

	Usual Hours	Actual Hours	Earnings
Self-report	34.15	32.85	\$314.86
Proxy report	32.89	32.45	\$303.88
Difference (F-value)	5.13*	.37	1.33
Correlation (Pearson r)	.70**	.69**	.80**

Note: Unless otherwise noted, for all Tables: ** $p < .01$ * $p < .05$ + $p < .10$

Table 3. Communication Mode and Self-Proxy Agreement on CPS questions.

% Using as a Communication Mode	Do Any Work for Pay		Doing Anything to Find Work	
	Self and Proxy Match		Self and Proxy Match	
	No	Yes	No	Yes
Conversation	42.9%	53.2%+	38.2%	50.6%
Learned from Someone else	12.2%	8.50%	8.8%	6.3%
Participation	18.4%	19.5%	17.7%	25.3%
Observation	32.7%	52.0%*	26.5%	35.4%
Read something	3.1%	8.3%+	2.9%	2.5%
General Knowledge	35.7%	53.6%**	14.7%	46.8%**

Table 4. Correlations of Proxy reports with Self-reports for Proxies using Different Communication Modes.

Communication Mode	Actual Hours		Earnings (log)	
	No	Yes	No	Yes
Conversation	.61	.77*	.78	.77
Learned from Someone else	.69	.72	.78	.55
Participation	.71	.56	.76	.95
Observation	.63	.77*	.75	.95+
Read something	.68	.87	.74	.93**
General Knowledge	.72	.60	.76	.87

Table 5. Self-Proxy Agreement for different Relationship Types.

Relationship Type	Do Any Work for Pay	Doing Anything to Find Work
	Self and Proxy Match	Self and Proxy Match
Spouse	85.1%	81.8%
Parent	89.1%	80.0%
Child	86.0%	83.3%
Sibling	78.2%	50.0%+
Other relative	75.0%+	56.3%+
Other nonrelative	73.8%+	80.0%

+ $p < .05$ one-tailed comparison with spouse

Table 6. Self-Proxy Agreement for different Relationship Characteristics and the Proxy's self-rated Knowledge about the Target.

	Do Any Work for Pay		Doing Anything to Find Work	
	Low	High	Low	High
Activities Done Together	82.5%	84.4%	64.0%	71.9%
Time Spent Together	83.7%	82.2%	72.9%	65.4%+
Self-rated Knowledge	82.8%	83.6%	66.7%	72.9%*

Table 7. Correlations of Proxy reports with Self-reports by Relationship Characteristics.

Relationship Type	Actual Hours	Earnings (log)
Spouse	.77	.77
Parent	.78	.86
Child	.55*	.51*
Sibling	.37*	.70
Other relative	.66	.81
Other nonrelative	.75	.74

** $p < .01$ * $p < .05$ for comparisons made with spouse

Table 8. Correlations of Proxy reports with Self-reports for different Relationship Characteristics.

Relationship Characteristics	Actual Hours		Earnings (log)	
	Low	High	Low	High
Activities Done Together	.60	.75*	.70	.86**
Time Spent Together	.68	.69	.76	.79+
Self-rated Knowledge	.69	.68	.71	.85**