

THE EFFECT OF QUESTIONNAIRE REDESIGN ON CONCEPTUAL PROBLEMS IN THE CURRENT POPULATION SURVEY

Elizabeth Martin, Bureau of the Census and Anne Polivka, Bureau of Labor Statistics
Elizabeth Martin, Bureau of the Census, CSMR, Washington, DC 20233-4700

One of the primary purposes of the questionnaire design and cognitive research conducted jointly by Bureau of Labor Statistics (BLS) and the Census Bureau since 1986 has been to identify and correct biases arising from variability in how respondents and interviewers interpret key survey concepts, such as "work," "looking for work," "job," and "business." One of the features of the Current Population Survey (CPS) instrument has been its reliance on respondents' intuitive understandings of ordinary words and phrases, without explaining the intended technical meaning of key concepts. For example, the primary question used to determine whether a person worked during the reference week is, "Did...do any work at all LAST WEEK, not counting work around the house?" The problem is that respondents may not know what they should report as "work." Some respondents do not realize the question is not meant to include volunteer work, school work, and housework, which are not covered under the official CPS definition of work. Others fail to report activities which should be reported, such as unpaid work done in connection with a farm or family business, or casual labor done for just a few hours, such as mowing lawns or babysitting. The question is worded to exclude "work around the house," which may discourage reporting of true work activities occurring at home. Another weakness of the current questionnaire is that there are no questions to elicit certain key facts, so the interviewer must rely on volunteered information. Currently, there are no explicit questions to determine whether anyone in the household had a farm or business, and if so whether any other family members did unpaid work for the farm or business. Rather, the interviewer is supposed to pick up this information during the interview, and probe as necessary. A final problem is that the reference period of the question is ambiguous; many respondents do not interpret "last week" as including Sunday to Saturday, as intended in the survey.

Over the years, questions have been raised about the classification of part-time or casual work, unpaid work in a family business or farm, and work in the underground economy, all of which may be underreported in CPS. Three primary changes to improve reporting of work activities were introduced in a redesigned CPS questionnaire. (The new questionnaire is being benchmarked in an overlap sample for 18 months and will become the official CPS questionnaire in 1994. See Rothgeb and Cohany, 1992, and Rothgeb et al., 1991, for discussions of major questionnaire revisions.) First, a question was added to determine if anyone in the household had a business or farm, and persons who reported no paid work were asked if they had done any unpaid work in the family farm or business. Second, the "work" question was revised to refer explicitly to pay and profit, the principal criteria for bona fide work activities, by asking, "LAST WEEK, did ... do ANY work for (either) pay (or profit)?" (The reference to profit was included if the respondent had reported a business or farm in the household.) The third revision was

to explicitly define the reference period in an introduction to the work series: "Now I am going to ask a few questions about work-related activities LAST WEEK. By last week I mean the week beginning on Sunday, (Date) and ending on Saturday, (Date)."

In this paper, we describe two types of diagnostic measures which were used in the CPS questionnaire design research to identify problems of respondent comprehension, and we examine the effects of the current and redesigned questionnaires on respondents' interpretation and reporting of work activities. The two types of measures are hypothetical vignettes, and direct probing questions. Households were randomly assigned to be asked either vignettes (1 in 10) or direct probing questions (9 in 10) as part of a respondent debriefing interview conducted immediately after completing the final CPS interview. The debriefing interviews were administered as part of the Phase II CATI test, which experimentally compared the current and the revised CPS questionnaire. Between July and October 1991, interviews were conducted in 3,800-6,000 households per month, with one of the two alternative questionnaire versions randomly assigned per household. Households were selected by random digit dialling and stayed in sample for 4 months, receiving the same questionnaire version throughout their tenure. Households were administered the debriefing interview only after completing the CPS interview in their final month in sample.

The hypothetical vignettes tested how respondents classified marginal or problematic situations, to measure their understanding of the work concept. For example, respondents were asked, "Earlier I asked you a question about working...Now I want you to tell me how you would answer that question for each of the persons in the following imaginary work situations ...Sam spent 2 hours last week, painting a friend's house and was given 20 dollars." Previous work has shown that respondents commonly do not classify the hypothetical situations consistently with CPS definitions. CPS respondents hold diverse interpretations of the meaning of "work." The most common response pattern was too broad (that is, respondents included as "work," activities which CPS does not), and the second most common pattern was much too restrictive (respondents excluded all marginal work activities, even those counted as work by CPS) (see Campanelli, Martin, and Creighton, 1989). The former pattern suggests the potential for overreporting bias, and the latter, underreporting. Analysis of patterns of association among responses to vignettes suggests that respondents vary along a dimension of inclusiveness which affects their classification of different situations as "work" (Martin, Campanelli, and Fay, 1991).

The vignettes require respondents to interpret situations which do not apply to them, and thus cannot in themselves support direct inferences about the effects of comprehension errors on reporting. In previous work, vignettes have been useful primarily for identifying

situations which are commonly misunderstood by respondents or interviewers or both (Campanelli, Rothgeb, and Martin, 1989; Esposito et al., 1991). Probing questions may permit more direct inferences about the amount of misreporting under different versions of a question. To measure certain unreported work activities, the debriefing interview included the question, "In addition to people who have regular jobs, we are also interested in people who may work only a few hours per week. LAST WEEK did ... do any work at all even for as little as one hour?" If valid, the results of probes such as this may permit direct measurement of bias under alternative versions of a questionnaire.

In this paper, we wish to address two main questions. First, did the questionnaire revisions improve reporting of work activities in the CPS? Second, do our alternative diagnostic measures provide consistent and meaningful information about comprehension errors which affect reporting of work activities? In the first section, we discuss results of the work vignettes, and in the second we analyze patterns of missed employment. The final section compares results from the alternative diagnostic measures.

1. Results of Hypothetical Work Vignettes

In the early stages of the CPS questionnaire design research, careful review of the questionnaire led us to expect that certain marginal work situations would commonly be misreported by respondents who may not interpret key terms in the ways intended by CPS (see Bureau of Labor Statistics, 1986; Martin, 1987). To explore that possibility, vignettes were designed to illustrate marginal work situations believed to be especially problematic. Few of the situations illustrated in the vignettes were thought to be particularly common in themselves, but the variety of marginal or ambiguous situations likely to be misreported could result cumulatively in a substantial amount of misclassification error.

The objective was not to measure the magnitude of misclassification error. That was the intent of the direct probing questions discussed below. The purpose of the vignettes in the early stages of the research was to identify potential problem areas which needed to be addressed by redesigning the questions or by improving interviewer training. In the analysis reported here, we use the vignettes to measure the effects of question wording on respondents' interpretations of key concepts and to examine whether the question revisions solved the problems identified earlier.

Table 1 compares the results of the work vignettes following the current and redesigned CPS questionnaire in the Phase II CATI test completed in October 1991. The introduction and wording of the vignettes were varied to reinforce the effects of the alternative wordings of the work question. (The work question was changed from, "Did...do any work at all LAST WEEK, not counting work around the house?" to "LAST WEEK, did...do ANY work for (either) pay (or profit)?") Respondents' classification of the vignettes was highly sensitive to the change in question wording. The wording revision drastically reduced the proportion of respondents who would improperly include volunteer work, and "work" carried out to help a family member (see vignettes 2 and 6) and, in general, appears to have reduced positive responses to all vignettes which do not involve payment. The new questionnaire appears to have made it more likely that respondents correctly include casual paid labor, such as described in vignettes 1 and 4.

The results in Table 1 suggest that the new questionnaire is leading respondents to be more inclusive in some situations, and more restrictive in others, in their classification of work activities.

To examine the effects of the questionnaire revision on the pattern of responses to the work vignettes, we fitted loglinear models to the 8-way cross-classification of questionnaire version (current versus redesigned) and all 7 work vignettes. We were interested in examining the effects of the questionnaire revision on respondents' conceptual structure, as indicated by patterns of association among vignettes. Although a larger number of cases might reveal a more complex response structure, these data suggest that the questionnaire revision affected the proportion of respondents saying "yes" to particular vignettes, but did not affect associations among them. The best-fitting model ($X^2 = 60.13$, $df = 70$, $p = .79$, from which all nonsignificant effects have been deleted in a stepwise manner) includes 10 (of a possible 21) pairwise associations among vignettes (1 x 2, 1 x 4, 2 x 3, 2 x 4, 2 x 6, 3 x 5, 3 x 6, 4 x 5, 5 x 7, 6 x 7, all positive except 1 x 2 and 6 x 7) and significant questionnaire effects on responses to vignettes 2, 4, 6, and 7.

Two summary indexes were computed based on responses to the 7 vignettes: (a) probability of a correct response, and (b) probability of a yes response. Table 2 shows that, averaged over all 7 vignettes, the probability of responding correctly in terms of the CPS definition is virtually unchanged (.67 in the current and .69 in the redesigned version). However, the average probability of "yes" is lower in the revised questionnaire (.53 in the current compared to .42 in the redesign; the standard error of the difference is .018). Thus, the questionnaire revision appears to have created a somewhat more restrictive concept of work, although for one type of marginal employment (casual labor for a few hours), the revised questionnaire is more inclusive. The questionnaire revision apparently did not communicate the intended meaning of "work" more clearly to respondents, since the probability of correctly classifying the vignettes is unchanged.

These findings are interesting in the light of the results of a previous analysis of similar vignette data from a 1988 study (Martin, Campanelli, and Fay, 1991). Rasch analysis of the structure of responses to a set of 5 work vignettes suggested that the underlying response dimension is inclusiveness rather than correctness. Respondents varied in the degree to which they held inclusive versus restrictive definitions of "work," but there was no evidence that respondents varied in terms of their tendency to answer correctly. We speculated then that this finding "may have implications for how questionnaire designers go about their work, because it may mean that efforts to increase respondent accuracy are naive. It may be that surveys can be designed to induce respondents to be more inclusive or more restrictive in their reporting, but not more accurate, in terms of a precise survey definition" (Martin, Campanelli, and Fay, 1991: 274-5). Results in Table 2 appear consistent with this speculation, although further analysis is needed to explore response dimensions in the 1991 data.

In some respects, the greater restrictiveness of the revised questionnaire is beneficial, since one probable effect is to reduce reporting of non-work activities, such as volunteer work and other helping activities (see vignettes 2 and 6). (Fewer mentions of nonwork activities may help explain interviewers' reports that they do not have to probe

as much with the revised work question, which also more frequently tends to be asked as worded by interviewers, and answered adequately by respondents; see Polivka and Rothgeb, 1992.) On the other hand, the same results also imply reduced reporting of certain categories of work which should be reported, including unpaid work in a family business (vignette 3) and work for commission (vignette 7). It is important to keep in mind that the revised questionnaire no longer relies upon volunteered reports of unpaid work in a family business, so it matters less that respondents have a poor understanding of how to report situations such as in those vignettes 3 and 5. The new questionnaire elicits reports of activities related to a family business by direct questions rather than relying on respondent interpretations of what should be reported as "work."

In sum, the results of the vignette analysis support the following expectations: (1) the proportion of persons reporting work activities during the reference week should be slightly reduced in the revised questionnaire, except that the addition of direct questions about unpaid work in a family business or farm should lead to improved reporting of those activities; (2) the questionnaire revision should lead to slightly better reporting of casual labor, such as work for a few hours (see vignette 4) or by students (see vignette 1); (3) the questionnaire revision may lead to reduced reporting of work for commission or work compensated by other than "pay."

In fact, evidence from the Phase II test shows there is virtually no difference between the two versions of the questionnaire in the proportion of persons reported as working during the reference week: 57.7 and 58.2 percent in the current and redesigned questionnaires, respectively, based on samples of 16,175 and 15,609.

There is evidence the revised questionnaire is eliciting a larger number of reports of certain types of work activities, including work in a person's own or a family business or farm. The proportion of workers who were classified as self-employed or as unpaid workers in a family business rose slightly but significantly from 12.9 percent in the current questionnaire to 14.8 percent in the redesigned questionnaire ($N=3,384$ and $3,244$, respectively; $X^2=4.6$, $df=1$, $p<.032$).

The revised questionnaire also appears to elicit more reports of work activities involving relatively few hours. The proportion of workers reporting 19 or fewer hours of work during the reference week is 10.0 percent in the revised questionnaire compared to 8.8 in the current ($N=9,028$ and $8,652$, respectively; $X^2=8.5$, $df=1$, $p<.003$). The difference is small, but is consistent with the expectation that the revised questionnaire picks up more casual labor, as implied by the vignette analysis.

The proportion of currently-enrolled students identified as also working is slightly higher in the revised questionnaire (45.4 and 48.8 percent in the current and redesigned questionnaires, respectively, with $N=930$ and $1,045$) but the difference is not statistically significant.

Finally, contrary to the expectation based on vignette analysis, there is no evidence that the revised questionnaire does a worse job of identifying work on commission. The proportion of persons at work during the reference week who reported in the debriefing interview they usually receive commissions is 11.4 percent for the current questionnaire and 10.4 percent for the redesign ($N=518$ and 568 , respectively; the difference in proportions is not

significant).

In the next section, we evaluate the results of the probe for missed employment involving just a few hours of work. We consider the effects of the questionnaire on the measure of missed work, as well as patterns of correlation between respondents' demographic characteristics and missed work.

2. Results of the Direct Probe for Missed Work Activities

As noted in the introduction, 9 out of 10 persons who reported no work activities during the reference week were probed in the debriefing interview to identify those who may have worked just a few hours but failed to report the activity in the main interview. It is important to note that the probe is likely to identify only one component of missed employment, casual labor. No probing questions were directed specifically toward other components of missed work, such as work in the underground economy, work done to set up or run a business or farm, or work for irregular or atypical forms of payment, such as payment in kind. The direct probe for missed work activities was asked about the first eligible person over 16 in each household. Persons were eligible if they reported no work activities during the reference week. Of those asked the work probe, 3.81 and 2.94 percent under the current and revised questionnaires, respectively, reported work activities in response to the probe. (The difference is not statistically significant.) These percentages represent the proportion of people classified as not working who should have been reported as working, assuming they reported valid work activities in response to the debriefing probe.¹ However, we are more interested in examining the universe of persons who were working during the reference week, and analyzing the proportion who were only identified as working when they responded positively to the debriefing probe. To create this measure of missed work, we selected a group of workers, comparable to the nonworkers who were asked the probing question about work activities. Since the probing question was asked about the first eligible person on each household roster who reported not working last week,² we also selected for analysis the first person on each household roster who reported working one or more hours the previous week.³ A maximum of 2 persons per household (one worker and one nonworker) would have been selected by these criteria. Only those households in which the debriefing interview was administered and direct probes were asked were included. (That is, the 1 of 10 households in which vignettes instead of direct probes were asked, and households not in their final month in sample, were excluded.) It is important to note that the "first eligible person" rule does not give each household member an equal or known probability of selection, which is influenced by the listing order of the household roster. In those households in which there was only one worker or one non-worker, there is no selection bias, but bias could be present for the 30 percent of households with more than one eligible non-worker and the 51 percent with more than one eligible worker.⁴ Typically, more senior and male members are listed first, and the chances of being selected as "first eligible person" vary according to the characteristics of both households and persons. Thus, the results presented below cannot be generalized to the total sample of persons included in the Phase II test. However, two factors lend the results some credence despite the limitations of the "sample." First, the analysis explicitly introduces demographic variables so their

possibly confounding effects are controlled. Second, the selection of "first eligible person" was identical for both versions of the questionnaire, so comparisons of results under current and redesigned questionnaires should not be affected. However, in the discussion which follows, we refer to "sample" persons to remind the reader that the cases included in the analysis do not represent a true probability sample.

Overall, 2.87 and 1.99 percent of "sample" persons were working but their work activities were missed under the current and redesigned versions of the questionnaire, respectively. (This difference is marginally significant: $X^2=3.25$, $df=1$, $p=.071$.) Tables 3a-d show the relationships between missed work and age, gender, race, and education, for each version of the questionnaire. These tables reveal striking differences among demographic groups in the proportion of "sample" persons whose work activities were not identified in the CPS interview proper. In general, patterns of missed work are quite similar under both versions of the questionnaire. Table 3a shows that the proportion of "sample" persons whose work activities were missed is very high among both the young (16-24) and old (post retirement, 65 and older) and is quite low during the middle years, 25 to 64. The redesigned questionnaire uniformly, and significantly, reduced the proportion of "sample" persons in all age categories whose work activities were missed. Since a feature of the new questionnaire is a greatly curtailed set of questions about work activities for retired persons, it is reassuring to find that the reduced number of questions did not lead a larger proportion of post-retirement workers to be misclassified as not working. On the other hand, an aim of the redesigned questionnaire was to reduce the amount of missed employment among youth, which has long been hypothesized as a source of bias in CPS (National Commission on Employment and Unemployment Statistics, 1979). Although the overall bias appears reduced in the new questionnaire, the association between missed work and age is unchanged.

Table 3b shows the association of gender with missed work, with women having higher rates of missed work than men. The gender bias appears smaller in the redesigned than in the current questionnaire, but the interaction effect is not statistically significant.

Table 3c shows work activities are more likely to be missed for Black and other race persons than for Whites. The race differential was not statistically significant in the current CPS, but it is under the revised version of the questionnaire. Although the interaction effect is not statistically significant, it appears that reporting of work activities may have worsened slightly for Black and other race persons under the new questionnaire, which is disturbing. Later analysis will show that this differential is entirely due to differences between questionnaire versions in the pattern of race differences for females.

Finally, Table 3d shows that education is negatively associated with missed work: the more years of education "sample" persons had, the less likely their work activities were to be missed. The relationship between education and missed work is the same in both versions of the questionnaire.

To examine the net and combined effects of questionnaire version and demographic characteristics upon missed work activities, loglinear models were fitted to the full 6-way cross-classification of Questionnaire Version X Race X

Gender X Education X Age X Missed Work. Due to the sparseness of the sample, our ability to test higher order interaction terms is quite limited. However, 4 models, presented in Table 4, are of interest in evaluating the effects of the questionnaire upon missed work activities. Models 1 and 2 both include the main effects of the demographic variables upon Missed Work; Model 1 includes the main effect of Questionnaire Version on Missed Work and Model 2 excludes it. Comparison of the goodness-of-fit for the two models indicates that dropping the Q x Miss term from Model 1 does not result in a significant loss of fit, and that the main effect of Questionnaire Version on Missed Work (net of the main effects of demographic variables) is not significant.

Model 3 further examines the effect of Questionnaire Version on Missed Work by fixing all main effect and interaction terms involving demographic variables and Missed Work, and estimating the Q x Miss main effect. (This model fits the full Age x Ed x Gender x Race x Miss cross-classification, thus including all 1-, 2-, 3-, 4-, and 5-way interaction terms involving these variables.) This model yields the parameter estimate of $-.1062$ (s.e. = $.0695$, $X^2=2.34$, $p<.13$) for the main effect of Questionnaire on Missed Work. This result implies that the new questionnaire has lower levels of missed work, net of all possibly confounding effects of demographic variables, but the net overall improvement due to the questionnaire is not quite statistically significant.

Finally, Model 4 includes main effects of all variables, and a 4-way interaction term involving Race, Gender, Questionnaire, and Missed Work. (The model is hierarchical, so inclusion of a 4-variable interaction implies the inclusion of all 1-, 2-, and 3-variable interaction terms involving these variables.) The 4-way interaction is marginally significant ($p<.10$) and occurs because the relationship between Race and Missed Work varies in the two versions of the questionnaire, but only for women. Separate analysis of women reveals that the race differential in missed work is significantly greater ($p<.04$) in the new questionnaire than in the current version, net of the main effects of demographic variables (analysis not shown).

3. Conclusions

The answer to the first question we set ourselves is that there appear to be modest improvements in reporting of casual labor in the revised questionnaire, and improvements as well in reporting of work in connection with a family business or farm. Overall, the proportion of persons whose work activities were missed was marginally reduced, although there are no differences between questionnaire versions in the proportion of the total sample reported as working. The small differences we find in this test may be larger when we switch from a completely telephone-based sample to the full CPS sample, since marginal employment may be more common in non-telephone households.

Analysis of the results of the probing question for missed work confirms biases in reporting of work which had hitherto been suspected but not documented. There are striking differences among demographic groups in rates of missed work, which are highest among the oldest and youngest age groups, women, Blacks and other race persons, and less educated persons. It was an objective of the questionnaire redesign effort not only to reduce overall bias, but to reduce differential reporting bias. We achieved marginal improvements within all groups, but the patterns

appear to remain largely unchanged in the revised questionnaire.

The diagnostic measures developed and refined in the CPS questionnaire redesign effort appear to yield consistent and useful information about the nature and extent of response biases present in the data. The vignettes provide information about particular problems of comprehension and interpretation and permit insights into the conceptual structure which underlies respondents' classifications of different situations. Analysis of vignette data permits comparisons of the relative restrictiveness or inclusiveness of respondents' interpretations of key survey concepts under different versions of a question. The probing question provides new information about the underreporting of work activities, permitting direct estimates of bias. Bias estimates such as those we have begun to develop in this effort give us a measure, lacking until now, by which to judge future redesign efforts in CPS.

NOTES

1. Positive responses to the probe were followed up with, "What kind of work did ... do?" and "Did ... get paid for the work?" Review of the verbatim responses suggest that 8 (of 58) and 4 (of 38) positive responses in current and redesigned questionnaires, respectively, clearly did not describe "work" as defined by CPS. Other, valid responses describe primarily casual labor (babysitting, mowing lawns, or yardwork), odd jobs or on-call work (security jobs, cleaning houses for pay, parking cars, construction and repair), farmwork or unpaid work on a family business, and work for pay which somehow went unreported, perhaps because it did not seem like "work" or a "job" (including teaching and tutoring for pay, therapy for pay, art, music and craftwork for pay). The analysis is based on all positive responses.
2. Of the households eligible for this debriefing question, 70.0 percent had only one nonworking member who was eligible, 26.6 had 2 eligible members, and 3.4 had more than 2.
3. Of the eligible households, 49.3 had only one eligible worker in the household and 50.7 had more than one.
4. In the overlap sample, which runs for 18 months from July 1992-December 1993, these debriefing probes will be administered for all eligible household members to eliminate this source of bias.

REFERENCES

Bureau of Labor Statistics (1986) Report of the BLS-Census Bureau Questionnaire Design Task Force. Washington, DC: BLS.

Campanelli, P. C., E. A. Martin, and K. P. Creighton (1989) Respondents' Understanding of Labor Force Concepts: Insights from Debriefing Studies. Proceedings of the Fifth Annual Research Conference, pp. 361-374. Washington, DC: Bureau of the Census.

Campanelli, P. C., J. M. Rothgeb, and E. A. Martin (1989) The Role of Respondent Comprehension and Interviewer Knowledge in CPS Labor Force Classification. Proceedings of the American Statistical Association (Survey Research Methods Section).

Esposito, J. L., P. C. Campanelli, J. Rothgeb, and A. E. Polivka (1991) Determining which Questions Are Best: Methodologies for Evaluating Survey Questions. Proceedings of the American Statistical Association (Survey Research Methods Section).

Martin, E. A. (1987) Some Conceptual Problems in the Current Population Survey. Proceedings of the American Statistical Association (Survey Research Methods Section).

Martin, E. A., P. C. Campanelli, and R. E. Fay (1991) An Application of Rasch Analysis to Questionnaire Design: Using Vignettes to Study the Meaning of "Work" in the Current Population Survey. The Statistician 40: 265-276.

National Commission on Employment and Unemployment Statistics (1979) Counting the Labor Force. Washington DC: GPO.

Polivka, A. E. and J. M. Rothgeb (1992) An Illustration of the Methods Used to Redesign the Current Population Survey Questionnaire. Presentation for the annual meeting of the American Association for Public Opinion Research, May 1992.

Rothgeb, J. M. and S. Cohany (1992) The Revised CPS Questionnaire: Differences between the Current and the Proposed Questionnaire. Paper prepared for presentation at the American Statistical Association meeting, August 1992.

Rothgeb, J. M., A. E. Polivka, K. P. Creighton, and S. R. Cohany (1991) Development of the Proposed Revised Current Population Survey. Proceedings of the American Statistical Association (Survey Research Methods Section).

Table 1

Results of Work Vignettes, for Current and Redesigned Questionnaire Versions
Vignettes describing work situations¹

Percent "yes"
Current Revised
q'aire q'aire

Vignettes describing work situations ¹	Current q'aire	Revised q'aire
1 Bill attended his college classes and got paid to tend bar for a fraternity party one night last week. (Current q'aire): Would you report him as working last week, not counting work around the house? (Revised q'aire): Would you report him as working for pay (or profit) last week?	78	85
2 Last week, Susan put in 20 hours of volunteer service at a local hospital.	36	4
3 Last week, Amy spent 20 hours at home doing the accounting for her husband's business. She did not receive a paycheck.	46	29
4 Sam spent 2 hours last week, painting a friend's house and was given 20 dollars.	61	71
5 Last week, Sarah cleaned and painted the room of her house in preparation for setting up an antique shop there.	47	42
6 Fred helped his daughter out by taking care of his grandson two days last week while the boy's mother worked.	13	2
7 Cathy works as a real estate agent for commissions. Last week she showed houses but didn't sign any contracts.	89	61
Total number asked work vignettes	297	304

¹ According to CPS criteria, correct answers are "Yes" to vignettes 1,3,4,5, and 7, and "No" to 2 and 6.

TABLE 2
Probability of Correct and "Yes" Responses for
7 Work Vignettes

	Current q'aire		Revised q'aire	
	Mean	s.d.	Mean	s.d.
Probability of correct response	.67	.18	.69	.19
Probability of "yes" response	.53	.24	.42	.20

Note: For each respondent, the number of correct (or "yes") responses is divided by the number of vignettes to which valid responses were given.

TABLE 3
Relationship between Demographic Characteristics and
Missed Work
By Questionnaire Version

Percent of "sample" persons with missed work activities

	Current CPS		Revised CPS	
	Percent	N	Percent	N
a. AGE				
16-24 years old	9.55	157	8.05	174
25-39	1.84	814	.99	809
40-64	1.47	954	.83	845
65 and older	14.89	94	10.47	86
	$X^2 = 83.6, df = 3, p < .0001$		$X^2 = 74.6, df = 3, p < .0001$	
b. GENDER				
Male	1.68	1371	1.47	1293
Female	5.40	648	3.06	621
	$X^2 = 21.9, df = 1, p < .0001$		$X^2 = 5.5, df = 1, p < .05$	
c. RACE				
White	2.77	1767	1.65	1692
Black/other	3.70	243	4.59	218
	$X^2 = .7, df = 1, n.s.$		$X^2 = 8.5, df = 1, p < .005$	
d. YEARS OF EDUCATION				
0-11 Years	6.34	205	5.99	217
12 Years	2.86	735	1.55	708
13-15 Years	2.63	532	1.74	516
16 or more Years	1.88	532	1.08	465
	$X^2 = 10.8, df = 3, p < .02$		$X^2 = 20.6, df = 3, p < .0001$	

TABLE 4
Loglinear Models Fitting the Cross-Classification of
Questionnaire Version, Gender, Race, Age, Education, and Missed Work

Model Terms of the model ¹	X ²	df	p
1 All main effects: Q x Miss, Ed x Miss, Age x Miss, Gender x Miss, Race x Miss	29.54	44	.95
2 Main effects, excluding Q'aire effect: Ed x Miss, Age x Miss, Gender x Miss, Race x Miss	30.02	45	.96
3 Main effect of Q'aire, and all interactions between demographic variables and Miss: Q x Miss, Race x Gender x Ed x Age x Miss	8.79	15	.89
4 Main effects for Age, Education, and one 4-way interaction: Age x Miss, Ed x Miss, Q x Race x Gender x Miss	24.73	40	.97

Note: Variables are:

Questionnaire Version (Q): 1 = Current, 2 = Redesigned

Years of Education (Ed): 1 = 0-11, 2 = 12, 3 = 13-15, 4 = 16 +

Race: 1 = White, 2 = Black and Other Race

Age: 1 = 16-24 2 = 25-39, 3 = 40-64, 4 = 65 +

Gender: 1 = Male, 2 = Female

Missed Work (Miss): 1 = Reported as working last week in CPS interview

2 = Not reported as working last week in CPS interview, but work activity reported in debriefing probe

¹ All models are constrained to fit the cross-classification of the 5 independent variables, Questionnaire x Education x Race x Gender x Age