

Comparing Results from Telephone Reinterview with Unmoderated, Online Cognitive Interviewing

William Mockovak and Robin Kaplan

Bureau of Labor Statistics, 2 Massachusetts Ave, N.E., Washington, DC 20212

Abstract:

Results from an unreconciled telephone reinterview were compared with results from an unmoderated, online cognitive interview to test questions designed to measure specific vocational preparation (SVP). Debriefings from a field test had indicated that the questions used to measure SVP led to some respondent confusion and possible misreporting. Revised SVP questions were tested using a telephone reinterview and unmoderated, online cognitive testing. A reinterview with 30 employers was conducted without reconciliation by two experienced interviewers (one conducting the interview, and one observing), and focused on identifying comprehension problems. Business establishments were selected to provide a wide range of industries, geography, and size classes. The fully scripted online cognitive interview relied on 11 participants talking aloud to determine how well the questions worked. With one exception, both pretesting approaches identified the same problematic questions and issues, but the reinterview provided additional valuable information about a more effective reordering of the SVP questions.

Key Words: Unmoderated cognitive interviewing, online cognitive interviewing, reinterview, cognitive interviewing

1. Introduction

This paper describes a small study conducted to refine a series of survey questions used to determine specific vocational preparation (SVP). SVP is defined as the amount of lapsed time required by a typical worker to learn the techniques, acquire the information, and develop the facility needed for average performance in a specific job-worker situation.¹ SVP is one of the components of worker characteristics found in the *Dictionary of Occupational Titles* (U.S. Department of Labor, 1991). The SVP concept has proven useful as a means of stratifying occupations) in terms of their level of required vocational preparation (there are nine SVP levels) and requires the following four pieces of information:

1. What is the minimum education required? If no minimum, must workers be able to read and write?
2. Is a professional certification, state or industry license, or other training required before the worker would be hired? How long does each (certification/license, training) take?
3. Is prior work experience required? If so, how much?
4. What post-employment training is required for the worker to perform proficiently (for example, OJT, mentoring, etc.)? How long does each type of training last?

In late 2013 and 2014, the Bureau of Labor Statistics (BLS) conducted a series of small field tests to determine if its National Compensation Survey (NCS) platform could be effectively used to implement the Occupational Requirements Survey (ORS), which will collect a variety of occupational data relevant to the Social Security Administration's (SSA) disability programs, including SVP (Ferguson 2013). The current testing provided the opportunity to learn more about the SVP questions and how they were interpreted and answered by respondents.

¹ <http://www.onetonline.org/help/online/svp>

Debriefings of interviewers (Field Economists/FEs) and observers in ORS field tests indicated that questions asking about different aspects of specific vocational preparation (SVP) led to some respondent confusion and possible misreporting. A summary of pretest data also indicated that SVP levels for some occupations were not closely aligned with SVP levels that had been computed in the past for similar occupations. This comparison did not necessarily mean the obtained SVP levels were wrong. However, interviewer debriefings also suggested that the order in which the specific vocational preparation (SVP) questions were asked affected some respondents' answers, as did the question wording, especially a question that asked about time to average performance once a worker was hired. With these concerns, revisions to the questions were suggested and additional testing planned.

Rather than wait to test a revised set of SVP questions in a future field test, the decision was made to conduct a small focused study to learn more about their performance. In addition, since changes to the SVP questions were being considered, and measuring educational attainment is a topic of interest in the federal statistical community, a variation of a question suggested by a federal government interagency working group that asked about certifications, licensing, and training was included in this study (Bielick et al., 2013). The key research objectives were to determine the following:

- Test the revised wording of questions measuring Specific Vocational Preparation (SVP) to see if the wording is clearly understood (compare results with question wording used in a previous field test).
- Identify interpretation problems and suggest additional revisions to the question stems and/or response options to improve respondent understanding or to improve the data collection process.
- Determine if the revised questions can be asked efficiently without adding burden to the interview.
- Determine if results from online, unmoderated cognitive testing provide supporting, or contradictory, results to the phone reinterview.

2. Method

This study consisted of two separate data collection components: (1) an unreconciled phone reinterview and (2) unmoderated, online cognitive interviewing.

2.1 Phone Reinterview

Thirty businesses that had participated in a previous field test, which had ended four months before (called Phase 3), were reinterviewed by telephone (4.7% of the total Phase 3 sample). Four experienced interviewers (Field Economists), who had collected data in previous field tests, worked in pairs (one collecting, one observing) to collect the data. This approach helped ensure that the highest quality data were obtained, responses were clearly understood, and that all potential problems were identified. If answers differed from those obtained in the preceding field test, interviewers did not reconcile any differences. However, they used any observed differences to ask probing questions about respondent understanding of the questions.

Since the number of jobs in each establishment for which data had previously been collected varied from two to eight, interviewers were asked to select only two jobs to reduce respondent burden and to allow time for a more in-depth discussion of the answers. Also, at least 25 percent of the jobs selected had lower levels of SVP since these were of most interest.

Business establishments were selected to provide a wide range of industries and size classes. Since the focus of the study was on question interpretation and understanding, convenience sampling with minimal nonresponse follow-up was used.

Interviewers talked to the respondent who had participated in the Phase 3 field test, which was typically someone from the human resources area or the owner (small businesses). Field Economists conducting the phone reinterviews were asked to use the wording developed for the scripted questions, but were given the option of changing the question order to see if that had an impact on collection.

Field Economists completed two debriefing forms for each business – a respondent debriefing form and an FE debriefing form – and also participated in a group debriefing at the conclusion of data collection for this study.

The reinterview was designed to be similar to a cognitive interview with unscripted, spontaneous probing (Willis, 1999), except that it was conducted over the telephone with two interviewers. All data were collected within four days.

2.2 Online Cognitive Interviewing

An unmoderated, scripted online test asked 11 respondents to complete a set of self-administered questions while thinking “out loud” and to respond to follow-up, scripted probes. The online testing was conducted using participants obtained through the online testing service.² Recruiting criteria asked for participants to report on a position that they supervised, or if they weren’t a supervisor, to report on their current job. Each online testing session was limited to 20 minutes, and a video was produced of the entire session.

The online test included the same questions as the reinterview. A very appealing characteristic of the online testers is that they have extensive experience in talking “out loud,” so they provide an extremely rich source of verbal data. Follow-up probes were scripted, as shown for the following SVP question:³

Figure 1: Example of Scripted Questions Used in Online Testing

6. Once hired, how much time is typically required for on-the-job training or mentoring before a new worker can perform the work satisfactorily?

Follow-up Questions:

- What does "perform the work satisfactorily" mean to you?
- If additional on-the-job training or mentoring is required and you haven't already described it, please do so.

3. Results

After the phone reinterview, the data collected in this study were compared to the data collected in Phase 3. Only raw data distributions were compared. No reconciliation was done.

The following tables show the SVP questions asked in the previous phase of ORS testing (Phase 3) and the version of the question asked in the current study.

² TryMyUI: www.trymyui.com

³ A SurveyMonkey instrument, accessible through TryMyUI, was used to script the cognitive interview.

Table 1a (Question 1). Wording Used in Phase 3 and Current Study – Minimum Educational Requirement

Wording Used in Phase 3	Wording Used in Current Study
“What degree is required for this job?”	“What minimum level of education is required for this job?”

3.1 Educational Requirement Question

As shown in the preceding table, the question in the current study was revised to focus on “the minimum level of education” required for a job. The resulting distributions are shown in Table 1b.

Table 1b. Comparison of Required Education Obtained in Phase 3 and Current Study

	Education			
	None	HS	Assoc	Bachelor
Phase 3	20	35	2	3
Current Study	10	41	4	5

*Fischer’s Exact Test $\chi^2(3) = 4.97, p = 0.174, LRX: 5.057 (p=0.1677)$

Although the differences were not statistically significant, the question wording used in this study led to 50 percent fewer responses of “no degree required” and more reports of a high school diploma or equivalent being the minimum requirement. In addition, more Associate’s and Bachelor’s degrees were mentioned. Debriefing reports from preceding field tests, including the most recent Phase 3 test, had noted that when some respondents heard the wording used in Phase 3, a common reaction was that many of them thought only of college degrees, so the identification of more high school degrees was viewed as an improvement in the data.

3.1.1 Observations from the Field Economists

The Field Economists reported that the version of the question used in this study worked well. Focusing on the minimum education required instead of asking about a degree seemed to result in more accurate answers and eliminated previously encountered confusion about high school degrees.

3.1.2 Observations from the Online Test

Results from the online test sample revealed no occurrences of confusion associated with the question wording used in this study. All participants found the wording straightforward and answered the question without hesitation.

3.2 Literacy Requirement Question

If respondents indicated that no degree was required for a job, they were asked whether there was a literacy requirement. Since this wording remained the same between Phase 3 and this study, it is not included in this analysis.

3.3 Question about Certification, License, or Training

Table 2 shows the wording that was compared for a question asking about required certification, licensing, or training (CLT).

Table 2 (Question 2). Wording Used in Phase 3 and Current Study – Certification, License, or Training Required Prior to Employment

	Wording Used in Phase 3	Wording Used in Current Study
A	“Is there any other training, license(s), or certification required for this job?”	“Is any type of professional certification, state or industry license, or training required for this job?”
B	“How long does this training usually take?”	“How long does it take to obtain the (certification/ license)? How long does the training last?”

As shown in Table 2A, fewer respondents in this study reported that a certification, license, or training (CLT) was required compared to Phase 3. This difference is statistically significant ($\chi^2(1) = 7.06, p = 0.008$). A total of 23 jobs in the Phase 3 testing and 10 jobs in the current study required either a certification, license, or training (CLT). Although the mean length of time to complete the CLT was higher in the current study, this difference was not significant, $t(31) = -0.50, p = 0.62$.

Table 2A – Is a certification, license, or training (CLT) required?

	CLT Required	
	No	Yes
Phase 3	37	23
Current Study	50	10

Table 2B – How long does it take to obtain the (certification/license/training)?

	CLT Length (days)	
	Mean	SD
Phase 3	260.9	453.8
Current Study	352.6	541.4

When answers were compared between Phase 3 and the current study, 16 clear discrepancies were found (out of 60 job comparisons). When substantial discrepancies occurred, the cause was usually the inclusion in Phase 3 of some sort of optional certification or some type of in-house or on-the-job training, which is not the intent of this question.

3.3.1 Observations from the Field Economists

The Field Economists (FEs) commented that although the question wording was clearer than that used in Phase 3, it still resulted in some confusion. For example, hearing the word “training” in the question led some respondents to think about, and want to talk about, either previous experience or on-the-job training. For example, one respondent wanted to include continuing education, which is not part of SVP. In addition, FEs reported that some clarification was required in slightly more half of the interviews, revealing reporting problems. FEs offered the following suggestions to enhance the flow and clarity of the interview:

- Separate the certification, license, and training components in the question. FEs reported that some respondents simply missed the part of the question asking about a certification or license, and they had to return to this question later. Because of this, one FE suggested moving this question to the

end of the SVP question series. As an aside, another FE who had tried this question order reported less confusion and better results. This suggestion was also supported by other FEs.

- Another FE suggested adding the term “pre-employment” to this question to distinguish it from questions asking about on-the-job training and previous experience.

3.3.2 Observations from the Online Test

In the online testing, if a certification, state or industry license, or training was clearly required for the job, there was no apparent confusion. However, when uncertainty existed, some respondents seemed to want to stretch the intent of the question to include things that they thought might qualify. For example, one respondent considered a CPR qualification from the American Heart Association as a “license;” another included a driver’s license as a requirement since a job required driving; one respondent mentioned that a certificate in paralegal studies exists but although she mentioned it, it was not specifically required for the job she described; and a couple of respondents mentioned job-specific, company tests that were required before a person would be hired, for example, a writing and grammar test for a writer (none of these should be counted).

There was also some confusion about the length of time required to obtain a certification or license. For example, a school teacher wasn’t sure if her answer should apply to her initial hiring or if it applied to recertification, which occurs every five years. She ended up giving the time for recertification, which was not the desired answer. Other respondents did not give estimates for the amount of time required to obtain a license or certification, saying only that one was required before the person would be hired. There was also variability in responding to the amount of time required, with some respondents giving ranges of time.

3.4 Prior Work Experience Question

Table 4 shows the wording used to determine prior work experience (PWE). This is a two-part question, with the requirement for prior work experience (yes/no) the first part, followed by a question asking about the amount of prior work experience that was required.

Table 4. Wording Used in Phase 3 and Current Study – Prior Work Experience

Wording Used in Phase 3	Wording Used in the Current Study
“Is there a specific amount of prior work experience required for this job?” (yes/no)	“Is a specific amount of prior work experience required for this job?” (yes/no)

As shown in Table 5A, there is only one minor difference in question wording, and there were no significant differences in the number of respondents who indicated that prior work experience (PWE) was required in Phase 3 vs. the current study. However, 18 of the 60 jobs had discrepancies of at least 30 days in the amount of prior work experience required in Phase 3 vs. the current study, even though the mean difference in prior work experience was not statistically different.

Table 5A. Is prior work experience (PWE) required?

	PWE Required	
	No	Yes
Phase 3	46	14
Current Study	44	15

Table 5B. – How much prior work experience is required?

	PWE Length (days)	
	Mean	SD
Phase 3	627.64	497.44
Current Study	677.93	405.06

In some cases, it was clear that respondents interpreted prior work experience (PWE) differently in Phase 3 vs. the current study. For instance, in Phase 3 the job of Crusher Operator/Front End Loader reported 6,000 hours of training required to reach journey level and additional forklift certification requirements. This counted as three years of prior work experience that were not included in the current study. Uncertainty as to whether this experience should be included as prior work experience, certification, or on-the-job training may have caused these discrepancies.

However, in other cases, higher levels of prior work experience were found in the current study compared to Phase 3. For instance, a Customer Service Manager position required one year of prior work experience in the current study, but only one month was reported in Phase 3. This discrepancy may have occurred because the interviewer in Phase 3 noted that the one month of prior work experience was taken from the job description. However, it's unclear whether the respondent directly provided this figure or not in Phase 3. In other comparisons -

- No prior work experience was reported for a Customer Service Rep job in Phase 3, but two years of prior work experience were reported in the current study.

The FE who collected the data in Phase 3 reported difficulties selecting an appropriate SOC code⁴ for the HVAC industry, which may have caused the discrepancy (somewhat different jobs may have been evaluated). Similarly, no prior work experience was reported in Phase 3 for a Fitness Concierge position, but two years of prior work experience were reported in the current study. Again, the Phase 3 FE had some difficulty finding the proper SOC code and felt the position was closer to an information clerk. In the current study, it was noted that sometimes people in this position have a certification to teach fitness classes, so it's possible two different positions were evaluated in each test.

- For a Bookkeeping job, the Phase 3 prior work experience was coded as “Don’t Know” and as five years in the current study. In both tests, FEs noted that the establishment is hiring a higher caliber of candidates than previously (CPAs), and although not a requirement, can require a higher amount of prior work experience than in the past.
- An Admin Assistant position had no required prior work experience in Phase 3, but two years in the current study. It's unclear why this difference occurred, but the interviewer in the current study noted it was a difficult interview.

3.4.1 *Observations from the Field Economists*

An issue that has come up in previous data collection is that respondents have commented that they are able to obtain workers with more prior experience than in the past because of generally poor economic conditions and employment opportunities. FEs have dealt with this issue by asking respondents to focus on what work experience is required, and not what could be obtained given changing job markets. Another area of

⁴ The 2010 Standard Occupational Classification (SOC) system is used by federal statistical agencies to classify workers into occupational categories for the purpose of collecting, calculating, or disseminating data. All workers are classified into one of 840 detailed occupations according to their occupational definition: <http://www.bls.gov/soc/>

confusion that sometimes arises is that respondents do not distinguish between work experience that is required before someone is hired, and a training period required after someone is hired.

Similar issues occurred in this test. FEs noted that some respondents wanted to talk about prior work experience when they were asked about certifications, licensing, and training. Therefore, some FEs thought that the question about prior work experience should be asked immediately after the question about minimum educational requirements, especially in lower skilled jobs where little education and/or experience are required. Other FEs concurred with this suggestion.

3.4.2 Observations from the Online Test

No issues were identified in the online testing. This question worked well. So in this case, the reinterview provided more information about possible difficulties with the question.

3.5 Time to Adequate Performance

Table 6 shows the wording used to assess time to adequate work performance once a worker was hired.

Table 6. Wording Used in Phase 3 and Current Study – Time to Adequate Performance

Wording Used in Phase 3	Wording Used in Current Study
“Once hired, how long must someone work in this job to be able to independently perform the essential job functions in an adequate manner?”	“Once hired, how long does someone typically have to work as a/an (occupation) in this job to meet your minimum expectations?”

As shown in Table 6, in Phase 3, respondents were asked how long it takes for an employee to perform in an adequate manner, whereas in the current study, respondents were asked how long it takes an employee to meet an employer’s minimum expectations. As shown in Table 7, the latter question resulted in a significantly greater number of days reported $t(118) = -2.57, p = .01$.

This question was the most problematic among the questions. A total of 24 out of the 60 jobs compared had a discrepancy of at least 30 days between Phase 3 and the current study. In most cases (19 of the 24), the question used in the current study resulted in a **higher** estimate of the time to meet minimum expectations as compared to the Phase 3 question.

For example, a common problem was the need to clarify that this question does *not* refer to a probation period. Confusion also stemmed from being unclear as to whether this question refers to a hiring manager’s preferences, the minimum qualifications, qualifications of the typical person who gets hired, or the length of time it takes to assess whether an employee is a good fit for the position.

Table 7. Time to Reach Adequate Performance (AP) or Meet Minimum Expectations (ME)

	Time to AP/ME (days)	
	Mean	SD
Phase 3	53.1	85.4
Current Study	99.9	112.5

An independent samples t-test was significant, $t(118) = -2.57, p = 0.01$. A paired sample t-test (matching the same job at Phase 3 and in the current study) was also significant, $t(59) = -2.89$,

$p = 0.006$, and the correlation between the two questions was small and non-significant, $r(60) = .19$, $p = 0.13$.

3.5.1 Observations from the Field Economists

FEs reported that this question did not work well in previous tests and although somewhat improved, the revised question still led to respondent confusion in the reinterview. Field Economists reported that when they asked this question, they often found themselves clarifying that the time spent post-employment would include completing OJT, mentoring, internal company training, etc. Therefore, they suggested that the question wording should specifically include this verbiage. Other examples of confusion were respondents who reported the length of the probationary period for the job, or reported that when they hired someone, they expected the person to “hit the ground running” from Day 1.

3.5.2 Observations from the Online Test

This question also did not work well in the online test. Participants revealed incorrect assumptions or misinterpretations in about 30 percent of the cases. For example, when asked to rephrase this question or explain its intent, participants said the following:

- “How much experience is enough for me to be comfortable hiring someone?”
- “Once you hire someone, what is their training period to see if they would work out?”
- “How long does it take you to figure out the person is a fit for the job or not?”
- “This question is asking how long we expect someone to work in the position once hired to meet our employment expectations, and our employment expectations are specific to a two-year commitment after training.”

3.6 SVP Calculations

Direct comparisons between SVP calculations for jobs interviewed during Phase 3 vs. during the current study are shown in Attachment 3. Twenty jobs had an SVP discrepancy of two or more between Phase 3 and the current study (there are nine SVP levels).

As shown in the following table, although there is a difference of 1.98 in the average SVP between the groups, this difference was not statistically significant.⁵

Table 8. Comparison of SVP Levels

	SVP	
	Mean	SD
Phase 3	4.2	4.35
Current Study	2.2	2.01

The correlation between SVP results in Phase 3 and the current study was moderate, $r(60) = 0.42$, $p < 0.001$.

4. Discussion

Prior to this study, the primary approach for evaluating questions being developed for use in the Occupational Requirements Survey was a series of small field tests, which relied heavily on respondent, interviewer, and observer debriefings, along with summaries of the data, to assess how well questions were

⁵ Independent samples t-test was non-significant: $t(118) = 0.47$, $p = .64$. Paired sample t-test (matching the same job in Phase 3 and in the Educational Requirements Test) was also non-significant: $t(59) = 0.62$, $p = .53$.

working. However, field tests, even small ones, require numerous resources, are expensive, and take time to complete and analyze. Therefore, when feedback from early tests suggested that the SVP questions were not working optimally, this small study was designed to provide more timely feedback using the less expensive approaches of reinterview and unmoderated, online cognitive testing. The primary evaluation approach used in this study was the reinterview, which was designed as a cognitive interview conducted using spontaneous probing by two highly skilled interviewers. The unmoderated, online cognitive testing was designed to provide supplemental information and relied exclusively on talking aloud with scripted, follow-up probes. Since the same researchers were involved in analyzing the results from the reinterview and the online testing, the potential for confirmation bias exists. However, we hoped to counter this possibility by having two researchers independently code and interpret the online results before discussing them together.

The reinterview in this study was not designed to generate measures of response bias or simple response variance (Morton et al. 2008, Biemer and Forsman 1992), which would have required larger samples and a reconciliation step in which inconsistent answers were reconciled with the respondent. Instead, the evaluation approaches used in this study relied heavily on qualitative findings and focused on identifying comprehension and communication problems associated with the SVP questions.

Although the reinterview generally worked well, there was evidence that in a very small number of cases, the same jobs were not being compared between Phase 3 and this study. The comparisons were based on 8-digit SOC codes, which is a very fine level of distinction, and there was some evidence of miscoding in Phase 3. The four month interval between interviews may also have led to some differences, although they are expected to be minimal since jobs don't tend to change that quickly.

Both the reinterview and the online test approaches provided useful, consistent feedback, which resulted in changes to the question wording and, most notably, changes in the recommended question order. The most problematic questions (training/certification/licensing and time to adequate performance) were the most problematic in both testing modes, and similar problems were identified. However, because the highly experienced phone reinterviewers had been given the freedom to ask the questions in a different order, they reported more success using an alternative order, which seemed to address some of the respondent confusion.

This type of feedback did not occur in the online test because only one order of questions was tested. Interestingly, when asked directly in the respondent debriefing, none of the 30 phone respondents reported that any of the questions were confusing. A general consensus based on both the reinterview and online cognitive interviews was that the set of revised questions used in the current study worked better than those used in previous testing, but that some problems still remained. Therefore, a revised set of questions with a different order was generated for use in future tests. The most effective question order after this test was deemed to be minimum education, prior work experience, post-employment training, and required certification, licenses, or training (required before a person would be hired).

Although the reinterviews and online cognitive testing were conducted concurrently in this study, a future strategy could be to complete the reinterview testing first and then follow up with online testing to explore key issues discovered in the phone reinterviews. For example, one of the key findings of the reinterview was that an alternative ordering of the questions led to better respondent comprehension. This alternative ordering could have been tested online to see if the same results were obtained. However, there is also value in using alternative methods concurrently because it's a way to check for consistency of results (Jick, 1979). Some key advantages of the online testing are that it's relatively inexpensive, and results can be produced very quickly compared to other pretesting approaches. Moreover, the video clips themselves serve as important documentation of problems that can then be shared with stakeholders or with field staff to illustrate conceptual problems that are occurring.

References

- Bielick, S.; Cronen, S.; Stone, C.; Montaquila, J.; Brock Roth, S.; Boivin, S. 2013. The Adult Training and Education Survey (ATES) Pilot Study, U.S. Department of Education, Active URL: <http://nces.ed.gov/pubs2013/2013190.pdf>.
- Biemer, P. & Forsman, G. 1992. On the Quality of Reinterview Data with Application to the Current Population Survey. *Journal of the American Statistical Association*, Vol. 87, No. 420, 915-923.
- Ferguson, G. 2013. Testing the Collection of Occupational Requirements Data. In *JSM Proceedings*, Government Statistics Section. Alexandria, VA: American Statistical Association. 2291-2305.
- Jick, T. D. 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, Vol. 24, No. 4, Qualitative Methodology (Dec., 1979), 602-611.
- Morton, J.E., Mullin, P., & Biemer, P. 2008. Using reinterview and reconciliation methods to design and evaluate survey questions. *Survey Research Methods*, 2 (2):75-82.
- U.S. Department of Labor. 1991. Dictionary of occupational titles (Rev. 4th Edition). Washington DC: U.S. Government Printing Office.
- Willis, G. 1999. Cognitive Interviewing: A “How To” Guide. Active URL: <http://www.uiowa.edu/~c07b209/interview.pdf>

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.