Providing Definitions and Reading Answer Choices: Experimental Findings from a Behavioral Health Survey

Donna Eisenhower[1], Stephen Immerwahr[2], Gregg Van Ryzin[2], and Bonnie Kerker[1]
NYC Department of Health and Mental Hygiene[1]
Baruch College Survey Unit[2]

Keywords: Measurement error, questionnaire instructions, providing definitions, answer categories

1. Introduction and Background

This paper reports on a series of cognitive tests and split-ballot experiments that test two practices often used in telephone surveys of health risk behaviors and status: (1) the impact of reading a clarifying definition of an ambiguous term as part of a question, and (2) the impact of reading the entire list of answer choices, rather than using a conditional reading of lists.

Assumptions about standards of practice would suggest that including a definition or reading the entire list of answer choices should result in higher quality data by reducing ambiguity, misunderstanding, and other sources of measurement error. However, these specific assumptions have not been rigorously tested. Such tests are important because there are costs associated with implementing these practices (e.g., increasing financial costs by lengthening the interview, irritating the respondent, break-offs due to the interview length, etc.); therefore, the benefits must be well understood to determine the overall gain in conducting a survey in this manner.

To test these assumptions, we used the 2005 New York City Community Health Survey (NYC-CHS), which is based in part on the Behavioral Risk Factor Surveillance System (BRFSS) and is an annual random digit dial (RDD) telephone survey. To better understand the impact of reading lists and definitions, we conducted cognitive pre-test interviews with forty respondents. These tests indicated differences in comprehension, recall, and reporting among test respondents. These cognitive interviews did not, however, enable us to determine whether the inclusion of definitions or changes in reading of lists would lead to different results in a population survey. To test the impact of these practices, a random sample of 3,087 respondents were randomly assigned to two experimental conditions for the first wave of the 2005 NYC-CHS. Chi-Square tests were conducted to test statistical significance at the alpha=.05 level. While this experimental design only tests for significant differences rather than the determination of which test condition provides the more valid measurement, a significant difference still does indicate under what conditions the level of the survey estimates themselves would differ.

This paper presents the findings for four tests of definitions and four tests of reading lists. The findings for each will be organized by the description of the test; cognitive pretest findings; general experimental findings; and a discussion of the findings.

2. Experimental Tests for Providing Definitions

Four experimental tests for providing or omitting definitions of ambiguous terms were included, namely definitions for medical care, physical activity, race for those of Hispanic origin, and household annual income.

Hypothesis #1: Providing definitions will significantly affect the distribution of answer choices.

2.1 Definition of Medical Care

Question: "Was there a time in the past 12 months when you needed medical care, but did not get it?"

2.1.1 Treatment

Version 1: Interviewer reads definition after question: "Medical care includes doctor's visits, tests, procedures, prescription medication, and hospitalizations."
Version 2 (No treatment): No additional information provided to respondent

2.1.2 Cognitive test findings

Cognitive testing indicated that different respondents included or excluded different aspects of care as "medical care" when answering this question. These included hospitalizations, tests and procedures, dental care and prescriptions; others restricted their answer to

doctor visits. We used the types of medical care mentioned to create the definition that was used in the experiment; therefore, we expected the inclusion of the definition to yield a higher proportion of individuals responding "yes" to this question.

2.1.3 Experimental findings

Those who were read the definition of medical care were less likely to report needing but not getting medical care. When read a definition, 11% said "yes," significantly lower than the 14% among those to whom the definition was not read ($\chi 2 = 8.21$, df = 2, p= .017).

2.1.4 Discussion of findings

The impact of reading the definition was significant and, yet, it was in an unexpected direction, with more respondents reporting they did not get care when they needed it when the definition was not read. One potential explanation is a "recency effect," whereby participants respond only to the last (most recent) item read – in this case "hospitalizations." Alternatively, respondents may have had a broader definition of medical care than the one read to them. This finding cautions us to pay careful attention to the definition itself and to test alternative formulations or reformulations.

2.2 Definition of Physical Activity

Question: During the past 30 days, other than your regular job, did you participate in any physical activities or exercise?

2.2.1 Treatment

Version 1: No treatment (no additional information provided to respondent)
Version 2: Interviewer reads definition after question: "Physical activities include such activities as running, calisthenics, golf, gardening, or walking."

2.2.2 Cognitive test findings

Some respondents, when asked to decompose their answer, thought of some activities they did not include when answering the question, and in some cases they changed their original answer. Some respondents also wondered if they should include walking to and from work. Therefore, reading the definition was thought to act as a prompt to be more inclusive and result in higher percentage of "yes" answers.

2.2.3 Experimental findings

Those who were read the definition did report a higher level of physical activity. When the definition was read, 74% said "yes" to participation in physical activity, significantly higher than the 62% who responded the same way without the definition ($\chi 2 = 46.9$, df = 2, p= .000).

2.2.4 Discussion of findings

Including a definition of "physical activity" which includes a range of activities from sports to gardening to walking primed the respondent to think back and include more activities than might otherwise is the case. Walking is part of many New Yorkers lives as they head to and from work, shopping, or running errands. When New York City respondents are directed to include walking, one can expect a higher level of physical activity.

2.3 Clarification of Race for Hispanics (among self-identified Hispanics/Latinos only)

Question: Which one or more of the following, if any, would you say is your race?

2.3.1 Treatment

Version 1: No treatment (no additional information provided to respondent)
Version 2: Interviewer reads definition before question: "Some people, aside from being Hispanic, also consider themselves to be a member of a racial group."

2.3.2 Cognitive test findings

Cognitive pretest findings, as well interview monitoring in the previous year, indicated that many individuals of Hispanic origin insist that their race is "Hispanic." In the NYC-CHS, similar to the U.S. Census, we first determine if the respondent is Hispanic or Latino, then we identify the specific origin or ancestry (Puerto Rican, Cuban, Dominican etc), and then we ask their race. If they insist that their race is "Hispanic" it is recorded as "none of the above." Hispanics comprise about 25% of the completed interviews.

2.3.3 Experimental findings

Hispanics were just as likely to stick with "none of the above" for the race question when the clarifying statement was read (46% in both versions) ($\chi 2 = 5.95$, df = 7, p= .545).

2.3.4 Discussion of findings

This difference may not have reached statistical significance because of the relatively smaller sample size of Hispanics. Nonetheless, cultural preferences and emerging norms may suggest changing the survey convention of recording race separately from ethnicity for Hispanics. If survey researchers persist in asking Hispanics if they are Hispanic and then asking them their race (or vice versa), a "none of the above" category should be retained for those who reject identifying their race as separate from ethnicity. The retention of this category is necessary to avoid the loss of information by coding these sentiments as "don't know" or "refused."

2.4 Household annual income

Question: "The next question is about your combined household income. Is your household's annual income from all sources less than (200% of the YEAR poverty level, calculated using both the number of adults 18 years of age or older and the number of children under age 18 in the household)?"

2.4.1 Treatment

Version 1: No treatment (no additional information provided to respondent)
Version 2: Interviewer reads definition: "By household income we mean the combined income from everyone living in the household including even roommates or those on disability income."

2.4.2 Cognitive test findings

When probed using cognitive techniques, some respondents forgot to include all sources of household income. Specific sources excluded in error were disability income of a spouse, social security income of a retired partner, and in a few cases failure to report any income other than one's own income. Including a definition was thought to reduce measurement error by increasing retrieval of all sources of household income.

2.4.3 Experimental findings

There was no statistically significant difference between the two versions for household income. This held for each of the income categories. However, those who were provided a definition were more likely to refuse to answer the question. When the definition was read, 7% refused, significantly higher than the 4% when it was not read ($\chi 2 = 11.16$, df = 3, p = .011).

2.4.4 Discussion of findings

While there was no significant difference, the definition has been retained on the annual survey because it better assures consistency of interpretation across respondents. The greater number of refusals when the definition was read may be the result of the emphasis on all members and sources of income. This may have lead people to be reluctant to disclose information. Or, a lack of knowledge of the income sources for other members of the household may have increased the refusal rate.

3. Experimental Tests for Reading Long Lists of Answer Choices

Each of these questions has a long list of answer choices with at least seven answer categories plus "refused" and "don't know." The long list of categories is often the result of trying to combine multiple questions with simple yes and no responses into one question. The BRFSS often uses a "Read if Necessary" instruction, while the CHS has a slightly different conditional reading. The CHS instruction is to read the first answer choice and if that isn't picked, to read the rest. When monitoring interviewers, however, we noticed an inconsistent application of this conditional instruction across a large number of interviewers. Questions on health care insurance, place of cigarette purchase, employment status, and birth control method other than a condom to prevent pregnancy were examined as part of this experiment to test the difference related to reading all of the answer categories or conditionally reading them.

Hypothesis #2: Reading all of the answer categories in a long list will result in a significant difference in the distribution of answer choices.

3.1 Health care insurance

Question: What type of health care insurance do you use to pay for your doctor or hospital bills? Is it insurance through...

Response categories:
a. Your employer
or
b. Other valid responses, namely:
02 Someone else's employer
03 A plan that you or someone else buys on your own
04 Medicare
05 Family Health Plus or Medicaid
06 The military, CHAMPUS, Tri-Care, or the VA

10 Some other source
88 NONE (Not read aloud)
c. Don't know or Refused (Not read aloud)

3.1.1 Treatment

Version 1: Interviewer reads entire list of responses
Version 2: Interviewer reads entire list of responses only if the respondent does not initially name "[my] employer" as their answer.

3.1.2 Cognitive test findings

There were instances where respondents changed their answer after hearing the whole list. For example, one respondent picked answer #1 but after hearing the list picked "some other source." Others found it irritating to listen to the long list of health care insurance types when the interviewer had already been told that the first choice read -- "[my] employer" -- was the answer the respondent wished to give.

3.1.3 Experimental findings

Those who were read the entire list were less likely to say their type of health care insurance was their employer. When read the entire list, 48% of the named their employer, compared to 51% of those who were read the first response and the remaining responses only if needed ($\chi 2 = 3.83$, df = 1, p = .050).

3.1.4 Discussion of findings

Fewer chose the first and most prevalent response category "[my] employer" when all categories were read. However, reading the long list does add time and costs and does irritate those with a clear choose of employer. A preferred solution would be to ask a separate yes/no question for each choice. While also long, it is less irritating to simply respond as yes or no to each brief question and more likely to be accurate.

3.2 Place of Cigarette Purchase (among current smokers only)

Question: Where did you get that carton/pack/single-loosie? Was it

Response categories:
a. From a gas station, deli, or other store in New York City vs.
b. Other valid responses, namely:
02 From another person or on the street in New York City
03 Outside New York City but in New York State
04 In a different state
05 Through the Internet or mail
06 Indian reservation
07 Outside the US
88 Other (SPECIFY)
99 Don't know or Refused (Not read aloud)

3.2.1 Treatment

Version 1: Interviewer reads entire list of responses
Version 2: Interviewer reads entire list of responses only if the respondent does not initially name "From a gas station, deli, or other store in New York City" as their answer.

3.2.2 Cognitive test findings

Many respondents picked the incorrect answer category during pre-testing unless they heard the entire list. Many also changed their answer to "Other" responses, with interviewers typing in the location or source of their most recent cigarette.

3.2.3 Experimental findings

There was a difference in the distribution of answer choices in the expected direction but the difference was not statistically significant: 81% chose this answer when read the list, and 75% chose the first answer category when read the first category and the remaining answers only if needed ($\chi 2 = 2.90$, df = 1, p= .089).

3.2.4 Discussion of findings

The lack of significant difference might be due to the small sample size of smokers, which in 2005 was only 19% of the adult population. Alternatively, if the majority truly purchased cigarettes from stores in New York City, the method would not impact the results. Or, perhaps the questionable legality of some of the answer choices discouraged their selection regardless of test version.

3.3 Employment Status

Question: Are you currently:

Response categories:
01 Employed for wages or salary
b. All other valid responses, namely:
02 Self-employed
03 Out of work for more than 1 year
04 Out of work for less than 1 year
05 A Homemaker
06 A Student
07 Retired

08 Unable to work
c. Don't know or Refused (Not read aloud)

3.3.1 Treatment

Version 1: Interviewer reads entire list of responses
Version 2: Interviewer reads first response category, "Employed for wages or salary," and only continues reading the list if this is not the respondent's answer.

3.3.2 Cognitive test findings

Some respondents changed their answer when they were probed for more detail and the question was reread.

3.3.3 Experimental findings

Those read the entire list were significantly less likely to be recorded as being employed for wages or salary. When read the entire list, 51% chose the first category of employed for wages or salary compared to 54% when it was not read ($\chi 2 = 4.12$, df = 1, p = .042). More respondents chose self-employed, student, and retired when the entire list was read.

3.3.4 Discussion of findings

There are a few answer choices that represent individuals who are employed. When interviewers do not read the list, they may record the first category when the person simply says employed, rather than probe for more detail. Similarly, if the answer is unclear, the first category may be chosen.

3.4 Birth Control Method (among those who had heterosexual sex in past 12 months)

Question: What method did you use?

Response categories:
a. No other method used
b. All other valid responses, namely:
02 Birth Control Pills
03 Shots [Not read: Depo-Provera]
04 Withdrawal
05 Rhythm Method
06 Foam, Jelly, Cream
07 IUD
08 Norplant
09 Patch/Ortha-Evra
10 Tubes Tied/Vasectomy [Sterilization]
11 Other Method
12 Not Fertile [Too Old, Can't Conceive, Pregnant Already, Just Delivered, Etc.]
c. Don't know or Refused (Not read aloud)

3.4.1 Treatment

Version 1: Interviewer reads entire list of responses
Version 2: Interviewer reads entire list of responses only if respondent does not provide an answer that can be easily coded into one of the answer categories.

3.4.2 Cognitive test findings

When the respondent's answer was unclear, "no other method" seemed to be chosen often. In addition, when interviewers did not read the list, respondents sometimes seemed embarrassed to admit that they could not remember the name or brand of the contraception and tended in that case to pick "no other method."

3.4.3 Experimental findings

Those who were read the entire list were significantly less likely to respond, “no other method.” When the entire list was read, “no other method” was chosen by 53% of respondents, compared to 61% when it was read only if needed ($\chi 2 = 10.37$, df = 1, p = .001).

3.4.4 Discussion of findings

Before this question was asked, respondents were asked whether a condom was used at last heterosexual sex within the past 12 months. The birth control question in the NYC-CHS represents a series of questions from the National Survey of Family Growth. In the NYC-CHS, these questions were combined into one, resulting in eleven answer categories. This was done due to space constraints, and this alternative is often thought to be better than no question at all. However, it would be better to ask multiple questions with "yes/no" response options for each. Given our time constraints, reading this long list does appear to increase accuracy, although at some cost of respondent burden. In addition, cognitive testing and monitoring of the interviews indicated some respondent embarrassment if they did not know what to call their form of birth control. This embarrassment would be avoided if the entire list was read and they could pick a choice.

4. Summary and Conclusions

While providing definitions for ambiguous terms and reading answer choices from long lists are thought to be preferred practices to minimize measurement error, previous research has not examined these assumptions. Experimental research is the preferred method to test the scientific rigor of practice. This is particularly

important when one of the practices may have detrimental consequences to a survey. In our survey, for example, we were concerned that reading long lists or definitions might dissuade participants from completing the survey; these practices also bring extra cost.

In this paper, a random sample of 3,087 respondents was randomly assigned to two experimental conditions. Two of four tests of providing definitions found statistically significant differences, namely, for medical care and physical activity. The results of the experimental findings also indicated some unexpected effects. For example, when the definition was provided for medical care, fewer respondents reported not getting needed medical care. Reading the more inclusive definition was expected to increase the proportion of individuals positively responding to this question, but this was not the case. This may be because respondents focused on the last item on the list, hospitalizations, which is a relatively rare event. Or, the results might be due to a limited definition. Future research should focus on the formulation and experimental testing of the definitions themselves. Plus, the conditions under which a type and style of definition makes a difference would also benefit from further testing.

Providing definitions may have other benefits, even if they do not affect the survey results. For example, the clarification offered for the race question for Hispanics, while not statistically significant, may have added value by putting respondents at ease. The inclusion of a definition for household income did not produce significant differences, but did help to distinguish individual from household income with more assurance.

Of the four examples of reading or not reading the list, three of four tests were statistically significant. The impact of reading a list may be influenced by many factors, including the length of the list itself, the required knowledge and recognition of the subject matter, and the percentage that truly belong in the first category. The added cost and irritation for those who have a clear answer in mind may indicate the need to ask multiple questions with short and simpler answer choices.

While these findings only indicated a level of statistical difference, future research needs to establish which condition is the more valid measure. Similarly, data users must evaluate if findings which have reached a level of statistical difference have practical significance and if sizeable differences which did not reach statistical significance still have practical significance. For both providing definitions to ambiguous terms and reading long lists of answer choices, cognitive and experimental testing can inform survey design decisions. Certainly, the survey research field will also benefit from further embedded experiments on these issues.