

WHEN TWO QUESTIONS ARE BETTER THAN ONE

Floyd Jackson Fowler, Jr., Gregory Fitzgerald, UMASS Boston, Paul Beatty, NCHS
Floyd Jackson Fowler, Jr., UMASS Boston, 100 Morrissey Blvd. Boston, MA 02125-3393

Key Words: Question wording, testing questions, cognitive testing.

INTRODUCTION

An important challenge for designers of survey research instruments is how to ask questions that involve complex definitions or concepts. When designing such questions, researchers have to address a number of issues, including:

- a. Which terms and concepts to define, and in how much detail to define them
- b. How to word questions and where to put the complex definitions in the question

In some cases, a researcher may decide that the cognitive burden of a particular question may be too great to be carried by a single question; in such cases, breaking a single question into two or more questions may be the best way to collect the needed data.

The research reported here stems from a study, sponsored by the National Center for Health Statistics, of alternative ways of dealing with such complex questions. In the course of the project, there were five instances where a single question and multiple-question design were compared to see which was the best way to achieve the same objectives. This paper reports on the results.

METHODS

The focus of the project was on 19 questions that were found in the National Health Interview Survey, or its supplements, or in the Behavioral Risk Factor Surveillance Survey (BRFSS). The 19 questions were chosen because they involved some term or concept that involved a complex definition. By virtue of the source from which they were chosen, the questions also qualified as having an important objective that someone thought was worth measuring.

Each of the questions was initially subjected to cognitive evaluation with 10 cognitive interviews. Asking respondents for narrative summaries of questions and descriptions of their answers, along with immediate follow-up probes, were the main ways in which information about question comprehension and the nature of answers was obtained.

Based on the initial round of cognitive testing, a set of revised questions was designed, and a second round of cognitive testing was carried out.

We then conducted a telephone survey in 156 households around the country. For each of the 19 test questions, we devised what we considered to be an alternative way of collecting the same information that

would address some of the problems identified in cognitive testing. A randomization procedure was established so that half the telephone respondents answered the original questions while the other half answered the alternative version. We also tape recorded 112 interviews (with respondent permission of course), and the tape recordings were behavior coded to identify the rates at which questions were read as worded, the number and kinds of probes that interviewers had to use in order to get adequate answers, and respondent requests for clarification.

This paper focuses on the five instances in which it was decided that the best way to achieve the question objectives was to ask more than 1 question.

RESULTS

Original Question: When riding in the back seat of a car, do you wear a seatbelt all of the time, most of the time, some of the time, once in a while, or never?

Alternative: In the past year, have you ever ridden in the back seat of a car?

When you are riding in the back set of a car, do you wear a seatbelt all of the time, most of the time, some of the time, once in a while, or never?

Cognitive evaluation showed that this is a classic example of a question that has an embedded, hidden assumption: that everyone rides in the back seat of a car. Everyone does not ride in the back seats of cars, and those that do not have trouble with this question.

There was little evidence from the behavior coding of any difficulty in administering the original question. However, the substantive results show that there is a marked effect of asking two questions (Table 1). Only 8% of respondents when asked the original question volunteered that they did not ride in back seats, whereas 20% of the respondents who were asked the explicit screener question said that. Moreover, the distributions make it clear that those who did not ride in the back seat were particularly likely to say "Never" in response to the original question. In fact, the original question is two questions in one. A good number of the respondents keyed on and answered the second question, whether or not they rode in the back seat, rather than the first question, the main question objective, which is the frequency with which they use seatbelts. It can be seen that the results are highly distorted because of the way the question was asked. Almost certainly the two-question approach produces better and more valid data.

Original question: What kind of place do you go to when you need routine medical care, such as a physical examination? Is it a doctor's office, clinic or health

center, hospital emergency room, hospital outpatient clinic, or some other place?

Alternative: People can get routine medical care in different places, including in a doctor's office, a clinic or health center, a hospital emergency room, or a hospital outpatient clinic. Do you have a place where you go for routine medical care such as a physical examination?

What kind of place do you go to most often for routine care such as a physical examination? Would you say a doctor's office, a clinic or health center, a hospital emergency room, a hospital outpatient clinic, or some other place?

Cognitive testing found that this question has essentially the same problem as the preceding one: It includes a hidden assumption that people have a place where they get routine medical care. That is not true. Those who do not have such a place have a difficult time knowing what to do with the question. When we look at the substantive answers, it turns out that only 1% of respondents volunteered that they did not have a place where they got routine care, but 19% said that when asked a direct question. Again, the resulting statistics are quite different, though the pattern is not quite so clear as in the preceding case.

The behavior coding also shows the tendency of people to interrupt when the long list of response alternative is read on the phone. In fact, because the people hearing the two-question version had already heard the list, they were particularly likely to interrupt the question before it had all been read. The long, complex set of response alternatives is a problem with both versions of the question; the "improved" alternative did not address that weakness in the design of this question (Table 2).

Original Question: This question is about automobile injuries, including injuries from crashes, burns, and any other kind of accident. Have you ever had an injury because of your driving?

Alternative: This question is about automobile injuries, including injuries from crashes, burns, and any other kind of accident. Have you every had an injury while you were in a car?

Were you ever the driver when you were injured?

Were you ever injured because of your driving?

In cognitive testing, it was found that the original question gave respondents too many things to think about. Essentially, it asks them to deal with three different concepts: being injured, being a driver, and being at fault. We found that respondents were frequently keying on one or two but not all three of the issues they were to consider. The idea behind the three-question series was to separate the issues, so they could deal with them one at a time.

This change again had a significant effect on the answers. The number of people being classified as

having had an injury because of their driving dropped from 8% to 2%. This change is quite consistent with the hypothesis that the question was improved; that people were saying "yes" to the original question without taking into account all three of the issues that they were to consider before they said "yes". The behavior coding also suggested that breaking the question into three improved the ability of interviewers to administer the question in a standardized way (Table 3).

For these first three examples, we believe that there is evidence that the multiple question approach significantly improved the quality of the data that resulted. In each case, there was a statistically significant change in the response distributions; the kind of changes that were observed were consistent with the notion that the cognitive problem identified have been addressed; solving the problem would produce changes in the direction and of the kind that were observed.

Original Question: Do you have at least one working smoke detector on each floor of your home? Include a finished basement or attic.

Alternative: For the next question, please include finished basements and finished attics as floors of your home. How many floors does you home or apartment have?

Do you have at least one working smoke detector on each of the floors of your home or apartment?

In cognitive testing, we found that the biggest challenge for this question was figuring out how many floors people had that should be taken into account. This was particularly complicated for those who lived in multi-unit structures. The idea behind the alternative was to break the question into two separate tasks. We thought that if we could get the respondents to first focus on the floor problem, and provide an accurate answer to that, the quality of data would be better to the smoke detector question.

When we compared the substantive answers, there is a tendency ($p < .13$) for respondents to the two-question series to be more likely to say that all the floors have smoke detectors. Based on cognitive testing, we think that may be a change in the right direction, as one of the most common errors people made was to include floors in their answers that were not part of their dwelling units.

On the other hand, the behavior coding indicates that these are really hard questions. In particular, look at the way the question about the number of floors stimulated activity. Based on this and interviewer feedback, it was quite clear that the problem of defining which floors count remains; it was simply concentrated in the screener question rather than embedded into the single original question. The revised question did not do much to clarify what was meant by a floor. In the end, the process of data collection was not improved by the revision, and we are not sure at all that the quality of the

data is better either (Table 4).

Original Question: Which of these best describes your employers' smoking policy for indoor public or common areas, such as lobbies, restrooms, and lunchrooms? Would you say, not allowed in any indoor common area, allowed in some indoor common areas, including designated smoking areas, or allowed in all common areas?

Alternative: At the place where you work, is smoking allowed anywhere in the building?

Think about all indoor public or common areas, including lobbies, restrooms, and lunchrooms. Which of these best describes your employers smoking policy in common areas? Smoking is not allowed in any indoor common area, smoking is allowed in some indoor common areas including designated smoking areas, or smoking is allowed in all indoor common areas.

A cognitive problem identified in testing occurs when no smoking is allowed anywhere. When that is the case, choosing an alternative that is a subset of the places smoking is not allowed (not allowed in any common areas) was a confusing task. The idea behind the two-question alternative was to sort out the people for whom there was no smoking allowed, so that the response alternatives made more sense.

When we compared the substantive answers, the finding was that there was no difference in the answers that resulted from the two approaches. When we look at the behavior coding, it is clear that the two-question strategy produced a better question-and-answer process (Table 5). In response to the original question, there were a lot of reading errors and interruptions. In addition, from simply an administration perspective, identifying people who worked in places where no smoking was allowed, which turned out to be the majority, enabled interviewers to skip a very complex question. So, from a process perspective, the two-question version was an improvement. However, in retrospect, it is understandable

why there was no substantive effect on the data. Although the question was confused and the response alternatives did not fit the "no smoking at all" situation very well, there really was no ambiguity about where to put an answer. Given the choices, there was no basis for someone who worked in a place where smoking was not allowed to get into the wrong category. Hence, even though it is a sloppy question, the substantive results were not affected.

DISCUSSION

Designers of survey instruments are usually concerned about minimizing the number of questions. However, combining multiple concepts or complex concepts into a single question usually is not a good approach. In the examples, at least 3 of the test questions almost certainly were improved by turning them into 2-or-3-question series; the resulting data were almost certainly better. For the 2 remaining questions, the process of administration was improved when the smoking at work question was turned into two questions, even though the data were not affected. The smoke alarm question probably was not much improved. However, the alternative was not better primarily because the revision did not address one of the fundamental flaws of the question: the ambiguous definition of what floors the respondents were supposed to consider in answering the question.

In conclusion, when asking questions that potentially involve complex concepts, there are a number of options about how to make questions better. One of the strong candidates for improving such survey questions is to divide the complicated questions into two or more simpler questions. Such a strategy is likely to improve the question-and-answer process and, in some cases, as we demonstrated here, it may also produce more valid data.

Table 1.

Original Question: When riding in the back seat of a car, do you wear a seatbelt all of the time, most of the time, some of the time, once in a while, or never?

Alternative Sequence: 1) In the past year, have you ever ridden in the back seat of a car?

2) When you are riding in the back seat of a car, do you wear a seatbelt all of the time, most of the time, some of the time, once in a while, or never?

SEAT BELT USE IN BACK SEATS QUESTIONS

	All of the Time	Most of the Time	Some of the Time	Once in a While	Never	Don't Ride in Backs Seat	Total
ORIGINAL QUESTION	30%	17%	13%	8%	24%	8%	100% (n=77)
ALTERNATIVE SEQUENCE	42%	16%	8%	10%	4%	20%	100% (n=79)

$\chi^2 = 19.23; p \leq .002$

BEHAVIOR CODING FOR BACK SEAT SEAT BELT QUESTIONS

QUESTION	Ns	(%) READING ERRORS	(%) INTERRUPT	(%) PROBES	(%) RESPONDENT ASKS FOR CLARIFICATION
ORIGINAL	60	1.6%	1(1.6%)	8.3%	3 (5%)
ALTERNATIVE 1	51	5.9%		3.9%	1 (2%)
ALTERNATIVE 2	40	7.5%	4 (10%)	2.5%	1 (2.5%)

Table 2.

Original question: What kind of place do you go to when you need routine medical care, such as a physical examination? Is it a doctor's office, clinic or health center, hospital emergency room, hospital outpatient clinic, or some other place?

Alternative Sequence: 1) People can get routine medical care in different places, including in a doctor's office, a clinic or health center, a hospital emergency room, or a hospital outpatient clinic. Do you have a place where you go for routine medical care such as a physical examination?

2) What kind of place do you go to most often for routine care such as a physical examination? Would you say a doctor's office, a clinic or health center, a hospital emergency room, a hospital outpatient clinic, or some other place?

WHERE PEOPLE GET ROUTINE MEDICAL CARE

	Doctor's Office	Clinic or Health Center	Hospital Emergency Room	Hospital Outpatient Clinic	Some Other Place	Doesn't get Routine Care	TOTAL
ORIGINAL QUESTION	72%	19%	1%	4%	3%	1%	100% (n=79)
ALTERNATIVE SEQUENCE	64%	14%		3%		19%	100% (n=77)

$\chi^2 = 16.65; p \leq .005$

BEHAVIOR CODING FOR QUESTIONS ABOUT WHERE PEOPLE GET ROUTINE MEDICAL CARE

QUESTION	N	(%) READING ERRORS	(%) INTERRUPT	(%) PROBES	(%) RESPONDENT ASKS FOR CLARIFICATION
ORIGINAL	51	7.8%	8 (15.7%)	5.9%	
ALTERNATIVE 1	61	18.8%	1 (1.6%)	11.5%	3 (4.9%)
ALTERNATIVE 2	48	18.8%	14(29.2%)	2.1%	2 (4.2%)

Table 3.

Original Question: This question is about automobile injuries, including injuries from crashes, burns, and any other kind of accident. Have you ever had an injury because of your driving?

Alternative Sequence: 1) This question is about automobile injuries, including injuries from crashes, burns, and any other kind of accident. Have you every had an injury while you were in a car?

2) Were you ever the driver when you were injured?

3) Were you ever injured because of your driving?

INJURY BECAUSE OF DRIVING

	YES	NO	TOTAL
ORIGINAL QUESTION	8%	92%	100% (n=79)
ALTERNATIVE SEQUENCE	2%	98%	100% (n=77)

$p \leq .05$

BEHAVIOR CODING FOR QUESTIONS ABOUT INJURY BECAUSE OF DRIVING

QUESTION	N	(%) READING ERRORS	(%) INTERRUPT	(%) PROBES	(%) RESPONDENT ASKS FOR CLARIFICATION
ORIGINAL	51	17.6%		3.9%	2 (3.9%)
ALTERNATIVE 1	61	11.5%	1 (1.6%)	0.0%	1 (1.6%)
ALTERNATIVE 2	21	9.5%		0.0%	1 (9.5%)
ALTERNATIVE 3	14	7.1%		7.1%	

Table 4.

Original Question: Do you have at least one working smoke detector on each floor of your home? Include a finished basement or attic.

Alternative Sequence: 1) For the next question, please include finished basements and finished attics as floors of your home. How many floors does you home or apartment have?

2) Do you have at least one working smoke detector on each of the floors of your home or apartment?

SMOKE DETECTORS ON EACH FLOOR

	YES	NO	TOTAL
ORIGINAL QUESTION	90%	10%	100% (n=79)
ALTERNATIVE SEQUENCE	96%	4%	100% (n=77)

$p \leq .13$

BEHAVIOR CODING FOR SMOKE DETECTORS' QUESTIONS

QUESTION	N	(%) READING ERRORS	(%) INTERRUPT	(%) PROBES	(%) RESPONDENT ASKS FOR CLARIFICATION
ORIGINAL	51	5.9%	12(23.5%)	19.6%	1 (2%)
ALTERNATIVE 1	61	9.8%		31.1%	11 (18%)
ALTERNATIVE 2	60	21.7%	1 (1.7%)	0.0%	

Table 5.

Original Question: Which of these best describes your employer's smoking policy for indoor public or common areas, such as lobbies, restrooms, and lunchrooms? Would you say, not allowed in any indoor common area, allowed in some indoor common areas, including designated smoking areas, or allowed in all common areas?

Alternative Sequence: 1) At the place where you work, is smoking allowed anywhere in the building?
 2) Think about all indoor public or common areas, including lobbies, restrooms, and lunchrooms. Which of these best describes your employer's smoking policy in common areas? Smoking is not allowed in any indoor common area, smoking is allowed in some indoor common areas, including designated smoking areas, or smoking is allowed in all indoor common areas.

RULES FOR SMOKING AT WORK

	Not Allowed in Any Common Areas	Allowed in Some Common Areas	Allowed in All Common Areas	No Outside Employer/Self-Employed	NA	TOTAL
ORIGINAL QUESTION	72%	10%	5%	10%	2%	100% (n=58)
ALTERNATIVE SEQUENCE	78%	11%	6%		6%	100% (n=54)

$p \leq .2$

BEHAVIOR CODING FOR QUESTIONS ABOUT RULES FOR SMOKING AT WORK

QUESTION	N	(%) READING ERRORS	(%) INTERRUPT	(%) PROBES	(%) RESPONDENT ASKS FOR CLARIFICATION
ORIGINAL	39	20.5%	10(25.6%)	17.9%	
ALTERNATIVE 1	45	6.7%		2.2%	1 (2.2%)
ALTERNATIVE 2	12	0.0%	1 (8.3%)	16.7%	3 (25%)