Journal of Official Statistics, Vol. 26, No. 4, 2010, pp. 633–650

Increasing Respondents' Use of Definitions in Web Surveys

Andy Peytchev[1], *Frederick G. Conrad*[2], *Mick P. Couper*[3], *and Roger Tourangeau*[4]

Survey respondents may misinterpret the questions they are asked, potentially undermining the accuracy of their answers. One way to reduce this risk is to make definitions of key question concepts available to the respondents. In the current study we compared two methods of making definitions available to web survey respondents – displaying the definition with the question text and displaying the definition when respondents roll the mouse over the relevant question terms. When definitions were always displayed they were consulted more than when they required a rollover request. The length of the definitions did not affect how frequently they were used under either method of display. Respondents who completed training items designed to encourage definition use actually requested definitions less often, suggesting that they may value minimal effort over improved understanding. We conclude that at least for small numbers of questions, providing definitions with the question is likely to be the more effective approach than rollovers or hyperlinks.

Key words: Definitions; user interface; respondent burden; question clarification; web surveys.

1. Introduction

Several studies suggest that survey respondents misinterpret questions with alarming frequency (e.g., Belson 1981; Conrad and Schober 2000; Suessbrick, Schober, and Conrad 2000). These comprehension problems can substantially reduce the accuracy of survey results (Schober and Conrad 1997; Conrad and Schober 2000; Schober, Conrad, and Fricker 2004). Respondents can differ from one another in how they interpret questions, affecting not only accuracy but also the variance of survey estimates.

Providing definitions in interviews can help communicate the researchers' intended meaning and improve accuracy of responses (e.g., Schober and Conrad 1997; Conrad and Schober 2000; Schober, Conrad, and Fricker 2004). The problem is that respondents may not request definitions as often as they need them. Respondents may not realize that ordinary words (e.g., "usually," "job," and "bedroom") may be used in nonstandard ways in surveys and thus fail to request definitions (Tourangeau et al. 2006). Alternatively, respondents may recognize the value of requesting a definition but find that

[1] RTI International, 3040 Cornwallis Rd., Research Triangle Park, NC 27709, U.S.A. Email: andrey@umich.edu. Andy Peytchev was at the University of Michigan at the time of this study.
[2] Institute for Social Research, 426 Thompson St., Ann Arbor, MI 48104, U.S.A. Email: fconrad@isr.umich.edu
[3] Institute for Social Research, 426 Thompson St., Ann Arbor, MI 48104, U.S.A. Email: mcouper@isr.umich.edu
[4] University of Maryland, 1218 LeFrak Hall, College Park, MD 20742, U.S.A. Email: rtourang@survey.umd.edu
Acknowledgments: This study is part of a program of research on web survey design and data quality funded by the National Science Foundation (Grant SES-0106222) to Roger Tourangeau, Mick Couper, Frederick Conrad, and Reginald Baker, and National Institutes of Health (Grants R01 HD041386-01A1 and R01 HD041386-04A1) to Roger Tourangeau, Mick Couper, and Frederick Conrad. We thank Market Strategies, Inc., for developing and deploying the survey.

the process of planning and making the request requires more effort than they are willing to expend. Moreover, they may be embarrassed to indicate to an interviewer that they do not know the technical meaning of everyday words. In contrast, web questionnaires are self-administered and probably do not cause such embarrassment. Kreuter, Presser, and Tourangeau (2008) have shown that web respondents answer sensitive questions more honestly than respondents who are interviewed, suggesting that social impediments to requesting definitions (i.e., acknowledging lack of knowledge) should be much reduced in web surveys. Thus we might expect more frequent use of definitions in web questionnaires than in interviews. Cognitive impediments to requesting definitions can also be reduced on the web relative to interviews by designing the clarification feature so that it requires very little effort, e.g., mouse rollovers (Conrad, Couper, Tourangeau, and Peytchev 2006).

Another approach is to provide definitions, by default, to all respondents rather than requiring respondents to request them. The main concern with this approach is that it does not assure that all respondents will *use* the definitions. Adding a substantial amount of explanatory text to each question may simply lead respondents to ignore such information. A given definition may be useful only to a few respondents, and those who do not need it may be more likely to ignore subsequent definitions, including some that might actually be useful to them. Thus, presenting too much information to respondents may actually backfire, reducing the use of definitions.

When developing questionnaires, web survey designers are confronted with various ways of delivering clarification to respondents. The two methods we consider are (1) letting respondents request definitions when they believe this might be helpful (respondent-initiated) and (2) presenting the definitions to all respondents by default (always provided). We have chosen these two approaches because they involve minimal respondent effort yet differ in the degree and type of respondent initiative that is required. The two approaches may each be effective but for different reasons.

1.1. *Respondent-Initiated Clarification*

Web questionnaires might be expected to promote the use of definitions by simplifying the request – typically a click or mouse rollover – and by removing any stigma created by the presence of an interviewer. However, the results to date are not encouraging. In one laboratory web survey (Conrad, Schober, and Coiner 2007, Experiment 1), respondents requested definitions relatively often only when they were explicitly told that definitions were essential to accurate responding; without this instruction, respondents requested definitions infrequently and considerably less often than needed. In that study, respondents could request definitions by clicking highlighted question text. In a "field" web survey, Conrad, Couper, Tourangeau, and Peytchev (2006) observed that respondents requested definitions on only 13.8% of the occasions on which they might have, almost certainly leading to some misinterpretation of the questions. Respondents were, however, more likely to request definitions when it required less effort. In their first study, Conrad et al. (2006) observed that requests for definitions were more frequent when getting a definition required only a single mouse click than when it required two or more clicks. In the second study, almost four times as many respondents (36.5%) requested definitions when this was possible with a mouse rollover than when the interface required one (8.9%) or two (6.5%) clicks. In both of these studies, when respondents did obtain definitions they seemed to read them, as evidenced by different patterns of answers when definitions were and were not accessed.

1.2. *Always Provided*

To request a definition with a mouse, respondents can move the mouse (or other pointing device) to the highlighted word(s) and press a button or, under a lower effort design, simply roll the mouse over the relevant question text. Even less effort is required if the definition is displayed with the question text. Depending on the exact design, respondents may need to move their eyes to a slightly different location on the page but this seems likely to be easier and more automatic than moving a mouse, and it may not be necessary if the definition is close enough to the question text. Displaying definitions in this way, all the time, may convey the intended meaning to some respondents who were unaware they were misinterpreting the question and so would not have explicitly requested a definition.

However, there are several concerns with this approach. First, respondents may treat the definitions as "fine print" and ignore them. They may not even notice the definitions because they are always present, i.e., they do not suddenly appear as with a rollover or click, yet at the same time are easy to distinguish visually from the essential question text and so easy to avoid reading. Alternatively, respondents may notice the definitions but perceive the entire body of the definition and question text as forming a particularly long question; this could promote break-offs (premature termination of the questionnaire) or satisficing (suboptimal responding) rather than promoting reading of the definitions.

Still another concern with presenting definitions along with the questions is that, although they are potentially useful for respondents who may otherwise misinterpret the question, the definitions may be redundant for respondents who correctly understand the question, wasting their time. Respondents may wonder about the researchers' motives for presenting so much unnecessary information; for example, they may infer that the researcher intends the questionnaire for nonnative speakers (see Yan 2005). In contrast when respondents request definitions, they almost surely recognize the value of definitions.

1.3. *Length and Relevance of Definitions*

In addition to varying whether or not respondent initiative was required to obtain a definition, we manipulated the length of the definitions to test the idea that respondents might be deterred from reading longer definitions that they did not request but might read longer definitions they had requested. Finally, we tried to vary the degree to which respondents believed the definitions were relevant to their task by having some respondents complete training items designed to stress the potential benefits of understanding the questions as intended.

2. Experimental Methods and Data Collection

A web survey on health and lifestyles was conducted in December, 2004, and January, 2005, as part of our methodological research program. The questionnaire was implemented and fielded by Market Strategies. Respondents were recruited through two sources: (1) email invitations were sent to a sample selected from the Survey Sampling International (SSI) Survey Spot panel, an opt-in web survey panel; as an incentive to participate, sampled panel members were offered entrance in a sweepstakes for one hundred cash prizes totaling $10,000; (2) pop-up invitations were incorporated into AOL web pages to lead potential respondents to a site ("AOL Opinion Place") where they could choose from numerous surveys in which to participate; as an incentive, they were offered American Airlines miles. In the SSI Survey Spot sample, 29,772 invitations were sent; 1,498 people responded and 1,361 completed the entire survey. In the AOL sample the number of people who saw the pop-up invitation is unknown. Neither of these approaches is designed to produce samples that are representative of the general population. Due to the volunteer nature of these recruitment methods, there are no selection probabilities and metrics on how well these probabilities are preserved (i.e., response rates) are inappropriate and misleading (Bethlehem and Stoop 2007). Rather, in Kish's (1987) terminology, the focus is on randomization rather than representation. That is, a particular feature of the design is experimentally manipulated, such as the display of definitions, and respondents are chosen at random for exposure to the treatment (definitions). All else being equal, which random assignment to groups helps assure, the differences in outcome between those who did and did not receive the treatment can be attributed to the treatment. Even if the sample does not represent all possible respondents we can be confident that in the current sample the effect of the difference between groups is due to the treatment (definitions).

Among the 2,719 participants who started the questionnaire, 2,481 reached the end. Among those who answered the demographic questions, 44.8% reported they were male and 55.2% female; 21.1% reported they were under 35 years of age, 39.8% 35 to 54 years old, and 39.1% 55 years or older; 4.6% Hispanic origin, 85.6% non-Hispanic White, and 9.8% other non-Hispanic; 23.7% a high school education or less, 40.0% some college or an associates degree, and 36.2% a college degree; 54.8% reported incomes less than $50,000 and 45.2% $50,000 or more. In addition, 12.6% reported that this was their first web survey, 40.7% that they had completed 2 to 15 surveys, and 46.7% that they had completed over 15 web surveys; overall, 82.2% reported using the Internet every day. More pertinent to the experiment, it is possible that as opt-in panel members the SSI and AOL respondents are not overly conscientious and may be less likely to pay attention and more likely to speed through the survey than respondents recruited with probability methods; if this is indeed the case, it might reduce their use of definitions to clarify questions.

The data about definition use were collected in the context of eight questions concerning the amount of food and nutrients each respondent consumed (these questions are displayed in Figure 1). The response options ran from "Much less than I should" to "Much more than I should." We chose the same eight questions used by Conrad et al. (2006), Study 2, to enable broad comparisons between the current and previous studies. Based on findings in Conrad et al. (2006) we used presentation methods that minimized respondent actions needed in requesting definitions and used only useful definitions that presented new or

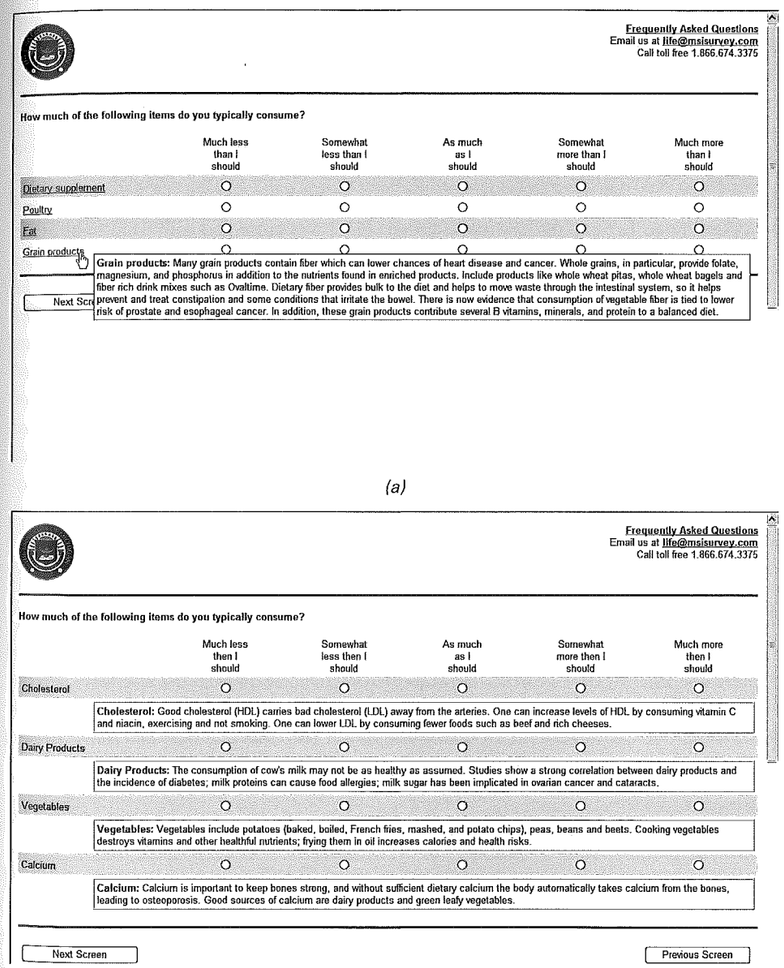

Fig. 1. First four questions in the rollover condition, with long definitions (a), and second set of four questions in the always-on condition, with short definitions (b)

counter-intuitive information, both of which increased definition use in the earlier study. We then varied (1) whether presentation of the definitions required respondent initiation, (2) the length of the definitions, and (3) whether respondents were "trained" in the benefit of definition use. While there were many other questions in this questionnaire, no definitions were available for them. The eight questions were presented on two pages (four per page) after 32 questions in a questionnaire with a maximum of 117 questions.

The order of the four questions on each page was randomized. The immediately preceding questions were on an unrelated topic, charitable giving.

2.1. *Obtaining Definitions*

For the critical questions, respondents were randomly assigned to one of three definition conditions: the control group (20% of the respondents) did not receive any definitions; the *rollover* group (40%) could obtain definitions by rolling the mouse over an underlined term (see Figure 1a); and the *always-on* group (40%) was presented with definitions at the same time as and immediately below each question (see Figure 1b). We assigned larger numbers of respondents to the latter two conditions in order to increase the power of the comparisons between methods of presenting definitions.

To underscore the utility of definitions (and to explain how they could be obtained in the rollover condition), the introductory page preceding the questions in the rollover and always-on conditions noted: "If you are uncertain about the meaning of a particular food or nutrient, please (hold the mouse pointer over the food or nutrient to obtain a definition/read the definitions for the foods or nutrients provided below the questions). Ordinary words can be used in unexpected ways in surveys so, to be sure you understand them, please (obtain/read the information provided for) more information."

2.2. *Length and Content of Definitions*

Half of the respondents in the rollover and always-on conditions were presented with short definitions (37 words on average), such as this definition of *cholesterol*:

> Good cholesterol (HDL) carries bad cholesterol (LDL) away from the arteries. One can increase levels of HDL by consuming vitamin C and niacin, exercising and not smoking. One can lower LDL by consuming fewer foods such as beef and rich cheeses.

The other half of respondents were presented with substantially longer definitions (126 words on average), such as:

> Good cholesterol (HDL) carries bad cholesterol (LDL) away from the arteries. One can increase levels of HDL by consuming vitamin C and niacin, exercising and not smoking. One can lower LDL by consuming fewer foods such as beef and rich cheeses. Cholesterol is a soft, waxy substance found among the lipids (fats) in the bloodstream and in all the body's cells. It's an important part of a healthy body because it's used to form cell membranes, some hormones and is needed for other functions. But a high level of cholesterol in the blood – hypercholesterolemia – is a major risk factor for coronary heart disease, which leads to heart attack. Cholesterol and other fats can't dissolve in the blood. They have to be transported to and from the cells by special carriers called lipoproteins.

The long version of each definition was designed to reiterate and expand the content of its shorter counterpart without introducing a substantial amount of new information (see the Appendix for the full set of definitions).

The definitions were constructed so that, if respondents read and considered them it would shift their answers relative to the answers of respondents who did not do so.

We achieved this by including information in the definitions that ran somewhat counter to common beliefs. For example, the definition for *cholesterol* stressed that there is not only bad cholesterol but also good cholesterol; similarly, the definition of *vegetables* indicated that vegetables can have adverse effects on health when fried. Although this provides relatively one-sided information, the direction of the effect on mean responses could not always be anticipated due to the complexity of deciding whether one eats more or less of something than one should. As a result, we used two-sided significance tests to assess the impact of the definitions.

2.3. *Training in Benefits of Definitions*

Prior to answering the eight experimental questions, half of the respondents in the rollover and always-on conditions were randomly presented with a page of "training questions." The training consisted of an assertion, "Many foods contain artificial sweeteners," followed by a definition of artificial sweeteners and two questions about the definition. One of the questions asked the respondent to report what was surprising about the definition and the other whether they had heard of some of the technical terms in the definition (see Figure 2). The overall purpose of the training was to raise respondents' awareness that without consulting definitions they might not understand the question as intended.

The experimental design consisted of eight experimental conditions created by crossing the method of obtaining definitions (rollover vs. always-on) with the length of the definitions (short vs. long) and presence of training questions (provided vs. not provided), with equal cell sizes; in addition, respondents in the control condition did not receive any definitions or training. Thus there were nine conditions altogether.

We recorded (as client-side paradata) whether or not respondents in the rollover condition requested definitions. In addition, at the end of the questionnaire we asked respondents whether they read any of the definitions ("Did you ever consult definitions

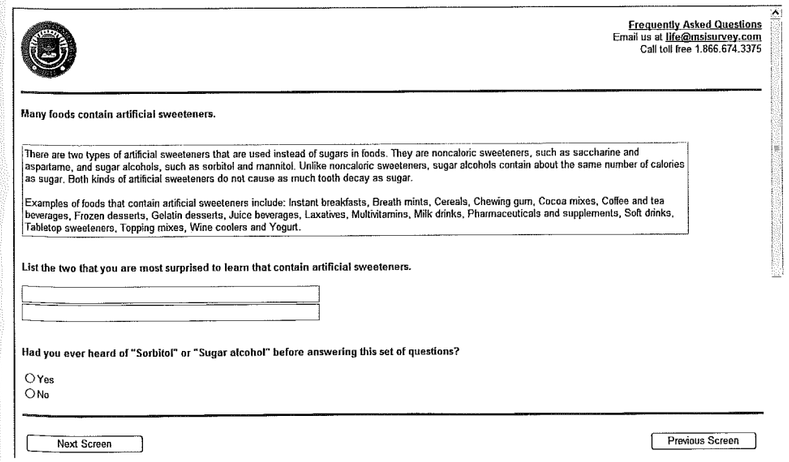

Fig. 2. Page with definition use training questions

that were available to you?") and recorded their answers. In all conditions, we captured the time spent on each screen.

3. Results

We examined how the different experimental treatments affected three main outcomes: use of the definitions, reading times, and distributions of responses. Our measures of definition use were based on self-reports and, where possible, automatically captured paradata. In addition, we also looked at breakoffs and item nonresponse because requesting or reading definitions might increase respondent burden, possibly leading respondents to abort the response task for the entire questionnaire (breakoff) or particular items (item nonresponse). We observed no significant differences in breakoff rates between conditions, possibly due to a floor effect. Similarly, we detected no differences in item nonresponse rates between any of the conditions.

3.1. *Definition Requests*

At the end of the questionnaire, respondents were asked several debriefing questions, including whether or not they ever used a definition. Almost twice as many respondents reported ever using definitions in the always-on condition, 60.7% ($n = 626$), than in the rollover condition, 35.6% ($n = 368$), further supporting the idea that the relative ease of moving one's eyes as compared to also moving one's hand leads to more definition use. We can be confident that, in general, self-reported definition use is accurate because respondents who reported having used a definition spent significantly more time on the two pages where definitions were available than respondents who did not, $F(1,2026) = 242.64$, $p < .001$.

We also recorded actual (as opposed to self-reported) rollovers in order to evaluate the self-reports from respondents in this group. These paradata were captured via client-side JavaScript. Because some rollover requests may not be requests at all but simply the result of moving the mouse into the sensitive area, the number of recorded rollovers is likely an overestimate of definition use. Furthermore, the question about definition use asked respondents whether they consulted the definitions, not whether they noticed them or glanced at them, so it is a conservative measure of definition use. Indeed, based on recorded definition requests, 45.4% of respondents in this condition rolled over at least one definition (Table 1), which is greater than the 35.6% who self-reported requesting

Table 1. Recorded rollover behavior by self-reported use of definitions in the rollover condition

Self-reported use of definitions	Rolled over	Did not roll over	Total
Yes	240	128	368
	65.2%	34.5%	100.0%
No	230	437	667
	34.5%	65.5%	100.0%
Total	470	565	1,035
	45.4%	54.6%	100.0%

Note: The percent of respondents who rolled over 1, 2, 3, 4, 5, 6, 7, and 8 definitions are 16.8%, 8.8%, 5.5%, 5.0%, 2.9%, 2.0%, 2.4%, and 2.2%, respectively.

definitions via rollover. Certainly some of these respondents intended to obtain a definition (they took 26 seconds longer than those who did not roll over once, $F(1,1060) = 94.71$, $p < .001$, suggesting that some of them consulted at least some definitions) but clearly not all rollovers were actually requests. Yet even this potentially inflated figure of rollover requests (45.4%) is smaller than the conservative figure (60.7%) of always-on respondents who reported consulting definitions. The overall message based on these measures is clear: substantially more respondents consult definitions when access requires only eye movements and not additional mouse movements.

The recorded rollover requests are also useful in examining the effects of the other two experimental variables – the use of training questions that emphasized the importance of definitions and the length of the definitions. We expected that respondents who were given the training questions would subsequently use the definitions at higher rates. We observed the opposite pattern – only 38% of those exposed to the training questions rolled over for any definition, while 53% of those who did not receive the training questions obtained at least one definition. For example, among those who got the training questions, 18% rolled over the definition for fat, but among those who did not get the training questions 29% did ($\chi^2(1) = 17.88$, $p < .001$). Respondents who are trained to weigh and consider the content of definitions may find this to be burdensome and so avoid requesting definitions when they are given the choice.

There was no impact of definition length on the number of respondents who obtained one or more definitions via rollover. Approximately 43% of respondents rolled over at least once in the short definitions condition and 48% did so in the long definitions ($\chi^2(1) = 2.52$, $p = .112$). The length of the definitions also did not affect the number of definitions respondents obtained in the rollover condition, looking just at those who rolled over at least once (Wilcoxon Mann-Whitney U test, $p = .986$).

3.2. *Response Distributions*

The answer categories running from "Much less than I should" to "Much more than I should" were recoded as 1 to 5 so that larger numbers reflect higher rates of consumption. For four of the eight questions (dietary supplements, poultry, fat, and cholesterol), answers varied across the three conditions (all Fs$(2,2652) \geq 4.59$, $p < .05$). As shown in Table 2, mean responses for two of these questions (poultry and fat) differed reliably from the no definition control in both the always-on and rollover conditions and by a larger margin for the always-on condition, although the difference between always-on and rollover for these two questions was not significant. For the other two questions (dietary supplements and cholesterol) mean responses differed significantly from those in the no definition group for only the always-on condition. The implication of these results is that respondents do read and consider definitions when answering survey questions at least some of the time. This seems to be especially likely when they can obtain the definitions by simply moving their eyes (always-on) as opposed to also moving their hands (rollover): the effect was observed for four of the items in Table 2 for the always-on condition but for two of the items for the rollover condition.

The length of the definitions did not have an effect on any of the consumption means. Coupled with the findings that the long and short definitions were not requested at different

Table 2. Mean responses (1 = Much less than I should, 5 = Much more than I should) to the eight questions and standard errors for each definition condition; significance levels reported for comparisons of each of the conditions in which definitions were available to the no definitions control group

	No definitions $n = 540$ Mean (S.E.)	Rollover $n = 1{,}082$ Mean (S.E.)	Always-on $n = 1{,}086$ Mean (S.E.)
Dietary supplements	3.40 (0.04)	3.45 (0.03)	3.27* (0.03)
Poultry	2.21 (0.04)	2.32* (0.03)	2.41*** (0.03)
Fat	2.74 (0.04)	2.64* (0.03)	2.60** (0.03)
Grain products	2.95 (0.03)	2.95 (0.02)	3.00 (0.03)
Cholesterol	2.53 (0.04)	2.49 (0.03)	2.68** (0.03)
Dairy products	2.66 (0.04)	2.67 (0.03)	2.64 (0.03)
Vegetables	3.21 (0.04)	3.30 (0.03)	3.12 (0.03)
Calcium	2.58 (0.04)	2.61 (0.02)	2.64 (0.03)

*Significant at $\alpha = .05$; **Significant at $\alpha = .01$; ***Significant at $\alpha = .001$.

rates but that availability of definitions in general led to different response distributions than no definitions, this suggests that if respondents did not read the longer definitions in their entirety they still gave them enough attention to affect mean responses.

The training affected answers only to the question on fat consumption. The mean answer on the five-point consumption scale for respondents who were administered the training was 2.57, compared to 2.66 for those who were not trained ($F(1,2121) = 6.06$, $p < .05$). The training was not differentially effective across definition presentation conditions for any of the questions including fat consumption. The lower fat consumption rates reported by respondents who received training presumably reflects the *relatively* positive description of fat in the definition. Nonetheless, respondents seemed insensitive to our efforts to increase awareness of the value of definitions when measured by patterns of answers.

3.3. *Time*

The response distribution provides a measure of the impact of a definition on subsequent answers but it does not by itself tell us much about the extent to which respondents read the definition. The time spent on each page provides additional detail about respondents' processing of definitions, since their response times should be longer when they read more text. To examine response times, we first compared the overall differences in time each respondent spent on the two pages adjusted for definition download time. By comparing times in the control condition to those in the rollover condition when a rollover did not occur, we can estimate the additional time needed to download the definitions in the always-on and rollover conditions. The difference in time between the fastest responses in the no definitions condition and the fastest times in the rollover condition among those who did not roll over any definitions yields an estimate of the time to download the additional text. This time was then subtracted from the times in each of the definition conditions, long and short. This adjustment process was carried out separately for the three different types of internet connections respondents reported using (i.e., broadband, dial-up, unknown). We also excluded 44 respondents who took more than five minutes on either page: 4, 11, and 29 cases in the no definitions, rollover, and always-on conditions,

respectively. Based on an ordinary least squares regression model, respondents in the rollover condition took seven seconds longer than their counterparts in the no definitions condition (mean = 62 seconds), $t(2,633) = 2.16$, $p < .05$, while respondents in the always-on condition took 60 seconds longer, $t(2,633) = 18.39$, $p < .001$.

Although respondents spent substantially more time in the always-on than in the rollover condition, the size of this difference could be driven by just the slowest readers for whom the extra reading disproportionately increased response time. As a check, we categorized respondents into ten response time groups (deciles) within each definition condition based on the total amount of time they spent on the eight questions and looked to see if the patterns were the same across the response time groups. Figure 3 shows that in all ten deciles respondents in the always-on condition spent more time on the eight questions than those in the rollover and control conditions (*t*-tests within all deciles significant at $\alpha = .05$). This strongly suggests that the respondents in the always-on condition – at all reading speeds – spent more time reading the definitions (or reading more words per definition) than their counterparts in the rollover condition.

In general, fast respondents (lower deciles) were less affected by the way definitions were presented than were slower respondents (higher deciles). For example, for the fastest respondents (first decile) mean response times were 28, 27, and 35 seconds for the no definition, rollover and always on conditions respectively; in contrast, for the slowest respondents (the tenth decile) mean response times were 134, 175, and 309 seconds for the no definition, rollover and always-on conditions. Clearly, slow readers were slowed more by the default display of the definition text than they were when the definition text was only displayed when requested via a rollover. It seems that because the fast respondents

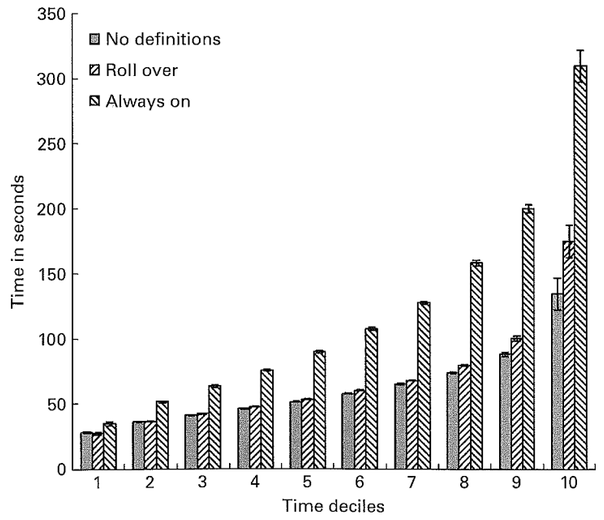

Fig. 3. Means and confidence intervals for time spent on the eight questions by definition presentation condition, in each decile

are fast readers (and some may have been less attentive), there is relatively little extra burden imposed by reading a definition but for slower readers the burden is disproportionately greater.

While respondents clearly spent more time answering and presumably reading the definitions in the always-on condition, it was not clear *a priori* whether this would be the case to the same extent for long definitions as for short ones. The results suggest that, overall, respondents spent more time answering questions when definitions were long than short, $F(1, 2100) = 12.89$, $p < .001$, and the effect of length was greater in the always-on (a 60-second difference) than in the rollover condition (a seven-second difference), producing a significant interaction of definition display and definition length, $F(1, 2100) = 6.03$, $p < .05$. It seems likely therefore that the always-on respondents read more total text than the rollover respondents but it is not clear that rollover respondents who requested definitions read less of the definitions than the always-on respondents do. To explore this we restricted the analysis of rollover respondents to just those who requested at least one definition; under these circumstances the interaction of definition display and definition length was not significant. We interpret this to mean that in the aggregate, respondents in the always-on condition consulted more of the definitions (read more text) than their rollover counterparts but when individual rollover respondents requested definitions they read them just as thoroughly as always-on respondents.

4. Summary and Conclusions

The evidence from this study indicates that when definitions are presented along with questions in web surveys, respondents are more likely to consult the definitions and consider their content than if a simple mouse action – merely a rollover – is required to obtain a definition. For survey designers, this means that it is more effective to display definitions together with the questions than to display them upon respondent request, even if the definitions are long. Mean responses differed from the no definition control for more questions in the always-on than in the rollover condition; respondents took longer to respond in the always-on than in the rollover condition; and more respondents reported using definitions in the always-on than in the rollover condition.

Despite a general advantage for the always-on method, when rollover respondents actually request definitions (as opposed to accidentally rolling the mouse over a sensitive region of the screen), they seem to read them at least as carefully as their always-on counterparts: definition length affects response times as much for rollover respondents who request definitions as for always-on respondents. When they were trained to find the valuable information in the definitions, rollover respondents were actually less likely to request definitions than when they were not trained, and there was no evidence that the training increased definition use for always-on respondents. Apparently, respondents who understand that the information contained in definitions may complicate the process of answering the question (e.g., the definition may require respondents to revise their understanding of a concept) try to avoid this extra effort. While respondents are sensitive to small differences in how much effort is required to use definitions, they do not seem to find the overall response task more burdensome when definitions are available than when

they are not: break-off and item nonresponse rates are no higher when definitions are available by either method than in the no definition control.

This study is partly an attempt to resolve a tension in the literature on clarification and definitions. On the one hand, reducing the effort needed to obtain definitions has been shown to increase the frequency with which they are requested (Conrad et al. 2006), but even when little effort is needed (a single click or just a mouse movement), few respondents access the definitions. On the other hand, providing definitions to all the respondents for all questions may push information on them that they do not need, making it less likely that they will read this kind of information when they do need it. The tension may be resolved by being clear about design goals. Is the goal to maximize the overall *use* of definitions or the *quality* of their use?

Based on the current data we have proposed that rollover respondents who request definitions spend as much time reading them as always-on respondents. In fact, rollover respondents who request definitions may read those definitions more thoroughly than always-on respondents. In a laboratory eye-tracking study, Galesic, Tourangeau, Couper, and Conrad (2008) observed that respondents were more likely to read at least some of the text in definitions that were always on than to request definitions via a rollover, but that those who did request a definition read significantly more words. Apparently respondents who are motivated to request a definition are also motivated to read it. By extension, embedding definitions immediately below the question text may invite somewhat more superficial processing by some respondents on some occasions, despite producing more total occasions of definition use. If one outcome – reading more definitions or reading definitions more completely – is more desirable than the other, then the design option is clear. Of course, designers may want the best of both worlds. A possible compromise is to present definitions along with questions but place the definitions before the response options and place the critical components of definitions in the beginning of the text.

While we believe our findings about definition use will generalize to other populations, the definitions used in this study were developed for experimental purposes. We do not wish to imply that these particular definitions should be used in surveys, but maintain that a good definition is one that presents information that not all respondents knew prior to participating. It was with this criterion in mind that these definitions were constructed. Our interest was in relative differences in definition use – which method is more effective. Similarly, our sample is not representative of the general population. We aimed to identify which method of providing definitions is more effective, but make no claim that the magnitude of the difference is a population estimate. We do note, however, that, some of these findings were replicated on a probability based national sample (Conrad et al. 2006).

Future research is needed to extend these findings to different survey conditions. Such variation might include the sampled population, the types of incentives offered to respondents, the questionnaire layout, the types of questions, and the definition content. We presented definitions for eight questions on two pages. Presenting definitions for all questions on all pages of a survey may change the effectiveness of always-on display, perhaps leading to better performance with respondent-requested definitions. While there may not be one optimal method for making definitions available under all designs, the intent behind survey questions is unclear often enough that web survey designers should routinely consider whether and how to clarify these intentions.

Appendix

Table A1. Definitions

Term	Short	Long
Grid 1		
Dairy products	The consumption of cow's milk may not be as healthy as assumed. Studies show a strong correlation between dairy products and the incidence of diabetes; milk proteins cause food allergies; and milk sugar has been implicated in ovarian cancer and cataracts.	The consumption of cow's milk may not be as healthy as has traditionally been assumed. For example, international studies now show a strong correlation between the use of dairy products and the incidence of diabetes; milk proteins are among the most common causes of food allergies; and milk sugar has been implicated in ovarian cancer and cataracts. The milk sugar lactose is broken down in the body into another sugar, galactose. In turn, galactose is broken down further by enzymes. According to a study, when dairy product consumption exceeds the enzymes' capacity to break down galactose, it can build up in the blood and may affect a woman's ovaries. Some women have particularly low levels of these enzymes, and when they consume dairy products on a regular basis, their risk of ovarian cancer can be triple that of other women.
Dietary supplement	Taking a dietary supplement, such as multivitamins, every day can improve your health in numerous ways. In particular, multivitamins can help protect cells against aging, improve sexual performance and reduce stress, among other benefits.	Taking a dietary supplement, such as multivitamins, every day can improve your health in numerous ways. In particular, multivitamins can help protect cells against aging (Vitamins C and E), improve sexual performance (Zinc and Vitamins C and E) and reduce stress (Omega-3 fatty acids, Valerian), among other benefits. A dietary supplement is a product (other than tobacco) that is intended to supplement what people eat; contains one or more dietary ingredients (including vitamins; minerals; herbs or other botanicals; amino acids; and other substances) or their constituents; is intended to be taken by mouth as a pill, capsule, tablet, or liquid; and is labeled on the front panel as being a dietary supplement.

Table A1. Continued

Term	Short	Long
Grain products	Many grain products contain fiber which can lower chances of heart disease and cancer. Products with whole grains, provide folate, magnesium, and phosphorus in addition to the nutrients found in enriched products.	Many grain products contain fiber which can lower chances of heart disease and cancer. Whole grains, in particular, provide folate, magnesium, and phosphorus in addition to the nutrients found in enriched products. Include whole grain bread and crackers, raisin bran cereal, and fiber rich drink mixes such as Ovaltine. Dietary fiber provides bulk to the diet and helps to move waste through the intestinal system, so it helps prevent and treat constipation and some conditions that irritate the bowel. There is now evidence that consumption of vegetable fiber is tied to lower risk of prostate and esophageal cancer. In addition, these grain products contribute several B vitamins, minerals, and protein to a balanced diet.
Poultry	Chicken, the most popular poultry bird, is less healthy than one might think. Pesticides and fungicides in chicken feed are passed to humans and could damage our central nervous system and cardiovascular system. Scientists have linked contaminated poultry to numerous cases of salmonella.	Chicken, the most popular poultry bird, is less healthy than one might think. Pesticides and fungicides in chicken feed are passed to humans and could damage our central nervous system and cardiovascular system. Scientists have linked contaminated poultry to numerous cases of salmonella each year. Eating chicken is generally no healthier than eating lean beef. Many people have the misconception that ground turkey is better (e.g., lower in fat) than ground beef. Ground turkey can be as high or higher in fat than ground beef because sometimes turkey is ground with dark meat and/or with skin. In addition, self-basting varieties of whole chickens or turkeys are higher in fat because they are injected with fat. Poultry includes chickens, ducks, geese, guinea fowl, peacock, pigeons, swans, and turkeys.
Grid 2		
Vegetables	Vegetables include potatoes (baked, boiled, French fries, mashed,	Vegetables include potatoes (baked, boiled, French fries, mashed, and potato chips), peas, beans and beets.

Table A1. Continued

Term	Short	Long
	and potato chips), peas, beans and beets. Cooking vegetables destroys vitamins and other healthful nutrients; frying them in oil increases calories and health risks.	Cooking vegetables destroys vitamins and other healthful nutrients; frying them in oil increases calories and health risks. Dark-green vegetables are: Beet greens, broccoli, collard greens, endive, escarole, kale, mustard greens, romaine lettuce, spinach, turnip greens, watercress. Examples of deep yellow vegetables are: Carrots, pumpkins, sweet potatoes, winter squash. Dry Beans and Peas (legumes) include: Black beans, black-eyed peas, chickpeas (garbanzos), kidney beans, lentils, lima beans (mature), mung beans, navy beans, pinto beans, split peas. Examples of starchy vegetables are: Corn, green peas, hominy, lima beans, potatoes, rutabaga. Examples of other vegetables include: Alfalfa sprouts, asparagus, bean sprouts, beets, Brussel sprouts, cabbage, cucumbers, eggplant, green beans, green peppers, lettuce, mushrooms, okra, onions (mature and green), radishes, summer squash, tomatoes, turnips, vegetable juices, zucchini.
Fat	Fat supplies essential fatty acids, which reduce chances for heart attacks, cancer, asthma, depression, accelerated aging, obesity, diabetes, arthritis, and Alzheimer's Disease. They are found in foods like canola and olive oil, nuts, fish, and chicken.	Fat is an essential part of the diet. It is the most concentrated source of energy we get, it supplies us with essential fatty acids, it promotes absorption of fat-soluble vitamins and it helps maintain healthy skin. Essential fatty acids also reduce chances for heart attacks, cancer, asthma, depression, accelerated aging, obesity, diabetes, arthritis, and Alzheimer's Disease. They are found in Flaxseed oil (flaxseed oil has the highest linolenic content of any food), flaxseeds, flaxseed meal, hempseed oil, hempseeds, walnuts, pumpkin seeds, Brazil nuts, sesame seeds, avocados, some dark leafy green vegetables (kale, spinach, purslane, mustard greens, collards, etc.), canola oil (cold-pressed and unrefined), soybean oil, wheat germ oil, salmon, mackerel, sardines,

Table A1. Continued

Term	Short	Long
		anchovies, albacore tuna, and others.
Cholesterol	Good cholesterol (HDL) carries bad cholesterol (LDL) away from the arteries. One can increase levels of HDL by consuming vitamin C and niacin, exercising and not smoking. One can lower LDL by consuming fewer foods such as beef and rich cheeses.	Good cholesterol (HDL) carries bad cholesterol (LDL) away from the arteries. One can increase levels of HDL by consuming vitamin C and niacin, exercising and not smoking. One can lower LDL by consuming fewer foods such as beef and rich cheeses. Cholesterol is a soft, waxy substance found among the lipids (fats) in the bloodstream and in all the body's cells. It's an important part of a healthy body because it's used to form cell membranes, some hormones and is needed for other functions. But a high level of cholesterol in the blood – hypercholesterolemia – is a major risk factor for coronary heart disease, which leads to heart attack. Cholesterol and other fats can't dissolve in the blood. They have to be transported to and from the cells by special carriers called lipoproteins.
Calcium	Calcium is important to keep bones strong, and without sufficient dietary calcium, the body automatically takes calcium from the bones, leading to osteoporosis. Good sources of calcium are dairy products and green leafy vegetables.	Calcium is important to keep bones strong, and without sufficient dietary calcium, the body automatically takes calcium from the bones, leading to osteoporosis. Good sources of calcium are dairy products and green leafy vegetables. In addition to reducing the incidence of osteoporosis, adequate amounts of calcium can: help control blood pressure, reduce the risk of colon cancer and help control weight. The best source of calcium is dairy products such as milk, yogurt and cheese. A single serving of cheese can give you 20 percent of the suggested daily intake. Other foods can provide calcium, too. Some examples include dark green vegetables, dried beans and calcium fortified juices and cereals. Dark green vegetables include beet greens, broccoli, collard greens, endive, escarole, kale, mustard greens, romaine lettuce, spinach, turnip greens, watercress.

5. References

Belson, W.A. (1981). The Design and Understanding of Survey Questions. Aldershot, UK: Gower.

Bethlehem, J. and Stoop, I. (2007). Online Panels – A Theft of a Paradigm? Proceedings of the Fifth International Conference of the Association of Survey Computing. Berkeley, UK: ASC.

Conrad, F.G., Couper, M.P., Tourangeau, R., and Peytchev, A. (2006). Use and Nonuse of Clarification Features in Web Surveys. Journal of Official Statistics, 22, 245–269.

Conrad, F.G. and Schober, M.F. (2000). Clarifying Question Meaning in a Household Telephone Survey. Public Opinion Quarterly, 64, 1–28.

Conrad, F.G., Schober, M.F., and Coiner, T. (2007). Bringing Features of Human Dialogue to Web Surveys. Applied Cognitive Psychology, 21, 165–188.

Galesic, M., Tourangeau, R., Couper, M.P., and Conrad, F.G. (2008). Eye-tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding. Public Opinion Quarterly, 72, 892–913.

Kish, L. (1987). Statistical Design for Research. New York: John Wiley.

Kreuter, F., Presser, S., and Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. Public Opinion Quarterly, 72, 847–865.

Schober, M.F. and Conrad, F.G. (1997). Does Conversational Interviewing Reduce Survey Measurement Error? Public Opinion Quarterly, 61, 576–602.

Schober, M.F., Conrad, F.G., and Fricker, S.S. (2004). Misunderstanding Standardized Language. Applied Cognitive Psychology, 18, 169–188.

Suessbrick, A.L., Schober, M.F., and Conrad, F.G. (2000). Different Respondents Interpret Ordinary Questions Quite Differently. Proceedings of the American Statistical Association, Section on Survey Research Methods, American Statistical Association.

Tourangeau, R., Conrad, F.G., Arens, Z., Fricker, S., Lee, S., and Smith, E. (2006). Everyday Concepts and Classification Errors: Judgments of Disability and Residence. Journal of Official Statistics, 22, 385–418.

Yan, T. (2005). Gricean Effects in Self-administered Surveys. PhD Dissertation at College Park, University of Maryland.

Received March 2009
Revised July 2010