Experiments in Mobile Web Survey Design

Andy Peytchev and Craig A. Hill
RTI International, 3040 Cornwallis Rd., Research Triangle Park, NC 27709

Abstract

Self-administered surveys can be conducted over mobile web capable devices, but the literature on mobile web survey design is scarce. Often, methods developed for an established mode of data collection are applied to a new mode. However, some established methods may be inappropriate. Mobile web surveys have unique features, such as administration on small screens and keyboards, different navigation, and reaching respondents in various situations – factors that can affect response processes.

Experiments were designed to address three main objectives. First, we test fundamental findings found robust across other modes, but whose impact may be diminished in mobile web surveys. Second, we test findings from experiments in (computer-administered) web surveys. Third, we experiment with the unique display, navigation, and input methods. While most findings from other modes are upheld, the small screen and keyboard on mobile devices introduce some undesirable differences in responses. Finally, we test attempts to alleviate these effects.

Key Words: Mobile devices, Cell phones, Smartphones, Survey design.

1. Introduction

Handheld and wireless technologies have advanced to the point where self administered surveys can now be conducted quite readily using Internet-capable mobile devices. Such devices have become increasingly widespread and their Internet capabilities widely used, with over 10% of mobile subscribers reporting being active mobile web users in countries such as the U.S., the U.K., Italy, and Spain (Nielsen Mobile, 2008). Typically, survey design in various modes has been informed through amassing systematic research, found absent for mobile web surveys. Modes of data collection for survey data differ not just in capabilities, but also in influences on survey responses. New modes for collection of survey data require research prior to implementation to minimize error, although implementation often precedes methodological work. Much can be borrowed from the research literature on other modes, yet much is likely different, and the erroneous, ill-considered or by-rote adoption of existing methods may have detrimental results. Research shows that surveys in each mode require unique considerations and designs, as demonstrated after the introduction of Computer Administered Telephone Interviewing (e.g., House and Nicholls II, 1988; Couper, 2000), Computer Assisted Personal Interviewing (e.g., Baker, Bradburn and Johnson, 1995; Couper, Hansen and Sadosky, 1997), and Computer Assisted Self Interviews (e.g., Ramos, Sedivi and Sweet, 1998; McCabe, 2004).

Mobile web surveys are displayed on devices that are substantially different from the devices now in common use by interviewers and respondents (i.e., laptops or desktop PCs), both in size and functionality (Jones et al., 1999; Watters, Duffy and Duffy, 2003; Chae and Kim, 2004; Parush and Yuviler-Gavish, 2004; Sweeney and Crestani, 2006), and can reach respondents in a variety of situations and locations (e.g., Brick et al., 2007). Systematic research is needed on mobile web surveys, rather than assuming applicability of methods for other modes.

There are two guided empirical approaches to developing methods for a new mode: replication of findings from other modes of data collection, and identification of likely differences that are then tested through new experiments. Replication of key findings from other modes is a relatively simple way of filling a methodological void for a new mode by linking it to a larger set of research. Different modes can share many factors that affect how questions are interpreted and how responses are edited through the same cognitive mechanisms (see Cannell, Miller and Oksenberg, 1981; Tourangeau, 1984; Strack and Martin, 1987).

However, replication of experiments within the context of other modes alone is not sufficient—it omits the effect of unique features in the new mode. For example, adopting self administered paper questionnaire designs for web surveys by placing multiple questions on scrolling pages (e.g., Dillman, 2000) can faithfully replicate mail survey experiments. However, doing so can limit the ability to reduce errors of omission through automatic skips (e.g., Couper et al., 1997; Peytchev, Couper, McCabe and Crawford, 2006), minimize item nonresponse through early validation (e.g., DeRouvray and Couper, 2002; Mooney, Rogers and Trunzo, 2003), and curtail possible correlated measurement error (e.g., Peytchev, 2007). Problems such as these can be ameliorated or eliminated if a paging survey design with one or few questions per page is used. Such findings illustrate that likely differences between modes need to be identified and tested to improve practices that could be suboptimal in the new mode.

Surveys can be presented using a browser-based Web application on an Internet-capable mobile device, just as is commonly done for computer-assisted self interviewing. Mobile devices differ from computer-administered web surveys in various ways, including: respondents can be in a larger variety of situations or locations that could affect cognitive processing; the devices have small displays that can limit the amount of information and affect how the survey is seen and comprehended; the methods of screen navigation and data entry are different, possibly affecting how the respondent interacts with the survey and what selections are made; and, programming functionality and application choices are often limited (thus far) compared to web surveys intended for administration on laptops or desktops. While we find absence of research of the effect of these differences on survey responses, related research literature on human-computer interaction and information processing finds that task success rates (such as correct selections) are lower on small screens (Jones et al., 1999; Parush and Yuviler-Gavish, 2004), and devices such as smartphones and PDAs with small screens and different navigation lead to less information gathering than when using computers (Sweeney and Crestani, 2006).

Mobile web surveys are possible and are being piloted in large scale studies (e.g., Okazaki, 2007), but are being conducted absent of methodological research on their design. It is possible that many design decisions can be imported from other modes of data collection. This new mode shares much of the functionality of web surveys and other computer assisted self interviews, such as the ability to present pictures and to present questions on separate pages. While much of the methodology for survey design in general, and web surveys in particular, would be relevant to mobile web surveys, there are nonetheless many unique characteristics. Compared

to paper instruments, data collection on mobile devices can be more efficient (Weber, Yarandi, Rowe and Weber, 2005), but is also reported by survey respondents as more difficult to use (Hardwick, Pulido and Adelson, 2007), which, if left unexplored, can be affecting data in unknown ways. Implementation of mobile web surveys requires systematic research that can start with testing similarities with other modes, but that also identifies likely differences.

2. General Design and Methods

In July 2007, we selected a probability sample and recruited adults into a mobile web survey panel with the purpose of studying this mode of data collection and evaluating the study protocol. A sample of 220 residential addresses was selected from two counties in North Carolina. Six interviewers made in-person contacts at each sampled dwelling unit, selecting the adult with the most recent birthday in each household. Participants were given smartphones and were provided with pre-paid voice and data. For their part, panel members signed a consent form on which they pledged to complete a short survey every week for a period of nine weeks. The panel was later extended to 19 weeks, continuing the pre-paid service. Ninety-two adults were selected and agreed to join the panel, a recruitment response rate of 43%. Note that this study design was not intended to produce estimates representative of the millions of mobile device users; rather, it provides a diverse set of respondents to whom methodological surveys could be presented on devices with known characteristics, allowing controlled experiments outside of a laboratory setting.

Panel members were invited to the 19 weekly surveys on topics such as politics, local issues, nutrition, and exercise via email or short message service (SMS). Surveys included about 15-20 questions, and cooperation rates to the weekly surveys reported here ranged between 59% and 70%.

Various design aspects of mobile web surveys were evaluated through ten experiments, embedded in four surveys in September and October 2007. We first present background and experiments for features that mobile web surveys likely share with other modes of data collection. We then turn to experiments targeted at likely commonalities with computer-administered web surveys. Finally, we describe rather unique features of mobile devices and present experiments that aim to evaluate the impact of these characteristics on survey responses.

3. Experiments on Mobile Web Survey Design

3.1 Design Aspects Shared with Multiple Modes

The survey response process model (Cannell et al., 1981; Tourangeau, 1984; Strack and Martin, 1987) proposes that respondents comprehend the question, retrieve information, make a judgment, and map their response to the available response options. This information processing also involves any editing, such as to conform to social norms, and has been labeled a process in its own right (Sudman, Bradburn and Schwarz, 1996). Such processes can be affected by survey design. For example, the ranges used in a frequency question about an undesirable behavior can be used by respondents to make an inference about the population distribution when responding, editing the reported frequency so as to conform to social norms. These effects of survey design on responses may be ubiquitous due to their general cognitive mechanisms, yet they may be moderated by mode.

Surveys on mobile devices could reach respondents in various locations and situations. It is possible that more haphazard responding occurs when the respondent is distracted or has less time, affecting cognitive effort in the response processes. Theories on dual processing would suggest that different information may be used if the respondent does not use a *central route* to processing, but rather a *peripheral route* (Petty and Cacioppo, 1986; Chaiken, 1987) that may be less thoughtful. That is, if individuals do not focus their attention as much when using a mobile device in some situations, less attention may be given to prior questions and responses to them, or less editing could be undertaken to conform to more socially acceptable behaviors.

This is a mechanism suggesting interaction between mode and response processing. To test it, key experiments in the literature from other modes were replicated.

3.1.1 Response Scales

Respondents use information in the response options. Even for a frequency question, they may alter their responses to avoid what would seem to be an extreme value for an undesirable behavior. This editing of responses can be conditional on the type of processing; so, for example, if respondents on mobile devices are engaging in peripheral processing, they may provide frequency reports that are unaffected by the response scale, counter to the findings in other modes of data collection (e.g., Schwarz et al., 1985; Tourangeau and Smith, 1996).

We conducted an experiment, randomly assigning panel members to a low frequency and a high frequency scale for a question asking the number of hours watching television, using the same wording as in a mail survey experiment by Schwarz and colleagues (1985), shown in Table 1.

Table 1. Responses to Low and High Frequency Scales to "How Many Hours a Day Do You Watch Television?"

Low Frequency Scale	n=28	High Frequency Scale	n=25
Up to ½ hour	4%	Up to 2 ½ hours	56%
½ to 1 hour	11%	2 ½ to 3 hours	24%
1 to 1 ½ hours	29%	3 to 3 ½ hours	4%
1 ½ to 2 hours	25%	3 ½ to 4 hours	0%
2 to 2 ½ hours	18%	4 to 4 ½ hours	8%
More than 2 ½ hours	14%	More than 4 ½ hours	8%

Results were consistent with findings in other modes. In our experiment, only 14% reported watching television more than 2 ½ hours in the low frequency condition, while 44% did so in the high frequency condition, a similar difference of 30% (χ^2, $p<.05$). Responses in the mobile web survey panel were substantially affected by the frequency ranges used in the scale, and the same considerations in the construction of frequency scales seem to apply. However, as discussed later, the possible solution suggested by Schwarz and colleagues of using open-ended questions in such circumstances may not apply to mobile web surveys if respondents are hesitant to use the small keyboard.

3.1.2 Question Order

Iran's authoritarian leader, Mahmoud Ahmadinejad, was invited to give a speech on September 24, 2007, at the University of Columbia at the peak of several controversies about his intentions of building nuclear capabilities, the treatment of women in Iran, funding terrorists in Iraq, and other issues. There were public and legislative outcries, condemning what was seen by some as providing a public platform for an authoritarian leader. Three weeks later, our panel members were asked: "Do you think that authoritarian leaders like Iran's President Mahmoud Ahmadinejad should be invited to give speeches at U.S. universities?" They were then asked: "Do you think that democratic leaders like President Bill Clinton should be allowed to give speeches at universities in authoritarian regimes like Iran?" For half of the panel members, the order of the questions was reversed, analogous to an experiment with questions about American reporters in the USSR and Communist reporters in the U.S. (Hyman and Sheatsley, 1950).

In the 1950 experiment, people were somewhat compelled to be even-handed and were found more agreeable to allow Communist reporters in the U.S. *after* saying that American reporters should be allowed in the USSR. If responding in mobile web surveys is less susceptible to earlier questions and responses, reporting about Ahmadinejad and Clinton should not be altered by order. Here again, results are consistent with findings in other modes, and the order of these questions significantly affected reporting. While 84% of respondents reported that they approved of Clinton speaking in authoritarian regimes when this question was presented first, only 57% approved when it came second (χ^2, $p<.05$)—closer to the approval of Ahmadinejad speaking in the U.S., as predicted by the norm of even-handedness. In both groups, 52% approved of President Ahmadinejad speaking in the U.S., consistent with the saliency of the issue at that time.

3.1.3 Fictitious Issues

An argument could be made that a survey is perceived differently when it is on a mobile device. At an extreme, should the respondent take the survey as seriously when it is displayed on their phone, as they would when administered by an interviewer or on a personal computer? Following Paul Grice's maxims on conversation (Grice, 1975), respondents infer that questions they are being asked are *relevant*, otherwise they would not have been asked. In other modes, this has been demonstrated as respondents provide attitudes towards fictitious issues (granted that there are alternative explanations for such responding).

In addition to evaluating whether respondents in mobile web surveys infer that questions must be relevant, it is also of interest whether responding to questions on fictitious issues can be affected by changing the framing of the task—a topic with import for survey design in general, as not all questions are necessarily relevant to all respondents.

Respondents were asked whether they favor or oppose the U.S.'s involvement in the war in Iraq, followed by two questions on fictitious issues, presented in Table 2. As shown in the last column, respondents voiced an opinion at noticeably lower rates for the fictitious conditions (43% and 59%) than for the nonfictitious question (83%), but, still, a substantially larger than zero proportion chose to appear knowledgeable about a fictitious issue. Thus, we once again find evidence that processing and inferences made about survey questions are fundamentally the same in mobile web surveys as in other modes: when respondents are given the survey task of answering questions, many will answer the questions as if it is pertinent to them.

Table 2. Responses to One Question on a Valid Issue and Two on Fictitious Issues, by whether a Textbox is Included with the Nonsubstantive Response Option.

	Without Textbox (n=29)	With Textbox (n=25)†	All (n=54)
1) Do you favor or oppose the United States' involvement in the war in Iraq?			
"Favor" or "Oppose"	79%	88%	83%
"Other/Don't Know"	21%	12%	17%
2) Do you favor or oppose the Intelligence Authorization Bill of 2008?			
"Favor" or "Oppose"	45%	42%	43%
"Other/Don't Know"	55%	58%	57%
3) Do you favor or oppose the Education Bill of 2007?			
"Favor" or "Oppose"	52%	68%	59%
"Other/Don't Know"	48%	32%	41%

† There was one missing response to the second question.

Research has examined whether respondents will provide attitudes about fictitious issues, but not about whether the underlying causes can be manipulated. If part of the cause stems from a respondent expectation that a researcher would not ask them something irrelevant or untrue, as suggested by Grice's *Logic and Conversation* (1975), then, if we could change that expectation, we should be able to alter the rate of "Don't Know" responses.

To test this, we conducted another experiment with a different set of questions on two fictitious issues and one valid issue, as shown in Table 3. In this experiment, respondents were randomly assigned to an introduction *encouraging* a response ("We really value your opinion on these current issues. Even if you have a weak opinion on this issue, it would be of help to us.") or *discouraging* it ("Some of these issues may not be well known. If you are not certain that you know about this issue, feel free to select 'Don't Know.'").

As in the first experiment, some respondents provided attitudes to the questions on fictitious issues, although very few did to the question on Lichtenstein, regardless of condition. Telling respondents that it is acceptable to not have an answer on these little-known issues significantly reduced opinions to the question about the Environmental Protection Act of 2007.[1] Respondents in this mobile web survey were not only susceptible to providing responses to fictitious issues as found in other modes, but the rate of such responses could be altered through manipulation of one likely mechanism.

[1] It is possible that some issues are interpreted as other, valid issues, but that itself is an error that survey researchers should work to minimize.

Table 3. Responses to Two Questions on Fictitious Issues and One on a Valid Issue, by Framing of the Question.

	Discouraging a Response" (n=27)	Encouraging a Response (n=27)†	All (n=54)
1) Do you favor or oppose the Environment Protection Act of 2007?			
"Favor" or "Oppose"	33%	63%*	48%
"Don't Know"	67%	37%	52%
2) Do you favor or oppose the recent political changes in Lichtenstein?			
"Favor" or "Oppose"	4%	15%	9%
"Don't Know"	96%	85%	91%
3) Do you favor or oppose Iran's development of nuclear facilities?			
"Favor" or "Oppose"	78%	92%	85%
"Don't Know"	22%	8%	15%

† There was one missing response to the third question; * p<.05

3.2 Design Aspects Shared with Computer-Administered Web Surveys

There are other design features that are not exclusive to, but are quite particular to, web surveys. Two such features are the ability to easily include images in the survey and the ability to vary the number of questions that are presented on a survey page. Both of these features have been found to affect responses (e.g., Couper, Traugott and Lamias, 2001). While mobile web surveys share these design capabilities, the effect on survey data may be very different. We further discuss and address each feature separately.

3.2.1 Ability to Present Images: Influence of Pictures on Question Comprehension

Whether images are used in web surveys in an attempt to make a more pleasing visual design, to assist retrieval of information, or for some other reason, images have been shown to affect responses to questions (Couper et al., 2004b; Witte et al., 2004; Couper et al., 2007). Indeed, pictures could be interpreted as part of the question, or alter question comprehension even as a contextual cue, depending on how they are noticed (see Couper, Tourangeau, Conrad and Crawford, 2004 on placement of the picture next to the question versus in the header of the page).

For images to affect responses, they must be noticed and processed. Like web surveys administered on a computer, mobile web surveys can display graphic information. However, with a small screen and other likely moderating situational factors, images may not impact responses in mobile web surveys. The size of the images did not significantly affect responses in an experiment in a computer-administered web survey (Couper, Conrad and Tourangeau, 2007), but the images in both conditions were relatively large compared to the size allowed on the screen of a mobile device. It could also be that images are more likely to be ignored on mobile devices, and while such factors may not be easily discernable in the present study, we were interested in whether images influence responses in mobile device surveys in the same way as they do in computer-administered web surveys.

We conducted an experiment to test whether images have such an impact in mobile web surveys, while considering that the effect of an image likely depends on its content and the question used. Participants were presented with a question on general health and randomly assigned to a picture of a person in a hospital bed or a person in competitive swimming, designed to alter the respondent's frame of reference, parallel to an experiment in web surveys (Couper et al., 2007). These same pictures were used with another question on physical fitness. Secondly, respondents were randomly assigned to pictures of people walking versus a person lifting weights while presented with questions on exercise; and, finally, respondents saw pictures of a sandwich versus fruits and vegetables while presented with questions on diet (question wording is presented in Table 4). The hypothesis, consistent with computer administered web surveys, is that the pictures will create a contrast effect such that respondents will respond somewhat relative to what is presented; i.e., people would report being healthier when in the context of a person in a hospital bed, as opposed to a person swimming at a sports event.

Table 4. Mean Responses in the Low and High Reference Picture Conditions.

Question text	Picture		
	Sick Woman (n=29)	*Swimmer* (n=30)	Sig. (t-test)
1) In general, how would you rate your health? (1=Very poor, 5=Very good)	4.0	4.1	0.69
	Walking (n=30)	*Lifting Weights* (n=27)	
2) In general, how often do you exercise? (1=Never, 5=Very frequently)	3.1	3.0	0.61
3) Compared to others, how often do you exercise? (1=Much less than average, 5=Much more than average)	2.9	2.6	0.27
4) Do you exercise three times a week for at least 20 minutes each time? (1=No, never, 4=Yes, every week)	2.2	2.2	0.97
	Sandwich (n=27)	*Vegetables* (n=31)	
5) In general, how often do you eat healthy food? (1=Never, 5=Very frequently)	3.5	4.0	0.04
6) Compared to others, how often do you eat healthy food? (1=Much less than average, 5=Much more than average)	3.4	3.4	0.92
7) Do you eat vegetables in at least one of your meals during the day? (1=No, never, 5=Yes, every day)	3.3	3.4	0.76

Under these conditions, we find limited support for images influencing responses in mobile web surveys. Mean responses were significantly different for only one of the seven questions and, indeed, the direction of the observed effects was counter to expectations—for example, the picture of vegetables seemed to lead respondents to report that they eat healthy food more frequently. This could be due to a less-than-typical image in the other condition (a *very* large sandwich). Furthermore, these mixed results illustrate that the effect of images in surveys can have rather unexpected results, as they are dependent on the *combination* of images and questions.

It is possible that situational factors determine how much attention is paid to the pictures. If so, we hypothesized that the differences may become significant if we account for the respondent's location. Inhibited by the limited sample size, we repeated the analysis controlling for self-reported location (at home, work, vehicle, or public location) using OLS regression. Results remained unchanged.

3.2.2 Ability to Separate Questions: Influence of the Number of Questions per Page

While web survey designers commonly have an option to decide how many questions to place on a page, there is very little guidance offered by the literature. Placing questions on the same page has been found to increase inter-item correlations (Fuchs, 2001; Tourangeau, Couper and Conrad, 2004), but others have found a decrease in such correlations (van Schaik and Ling, 2007). Modeling approaches suggest that the higher correlations are due to correlated measurement error (Peytchev, 2007).

The same design choices are present in mobile web surveys, but with an unknown effect on responses, if any. To attend to this question, respondents were randomly assigned to answer ten questions on activities of daily living—either one question on a page, or all ten on the same page. If similar to other web surveys, placing the questions in a grid on the same page should yield a higher proportion of respondents selecting the same response option, and higher inter-item correlations (as measured by a coefficient of internal consistency).

We did not find any differences in selecting the same response option (40% when separate versus 46% in a grid, p=.621) or in internal consistency (Cronbach's Alpha of .90 in both conditions); furthermore, we found no differences in means to each question across the conditions (data not shown). However, the lack of differences could also be attributable to the items used which were measured on three-point scales and had very limited variability. It would be beneficial to replicate this experiment with different items that have greater variation and on a larger sample in order to have sufficient power to detect differences in second-order statistics.

3.3 Design Aspects Particular to Mobile Web Surveys

Aside from commonalities with other modes of data collection, and with computer administered web surveys in particular, mobile web surveys are likely to exhibit unique measurement differences, whether due to different features or common features that are far more pronounced. These differences include the display of information, how the respondent navigates through the survey, and how responses are selected, in addition to external factors such as surroundings and connection problems. First, mobile devices have small screens, affecting the size of text and images, and the amount of information of information that can be displayed at any one time. An argument could be made that the same amount of information can indeed be placed on a mobile device's screen if this information is allowed to extend beyond the visible area on the screen. This information could only be seen, then, with additional actions by the user, but it is unknown whether respondents would be uniformly willing and able to perform the additional actions required. For an image, it is likely important that respondents see the image in its entirety, yet reducing its size to achieve this may have the effects discussed in the previous section.

Second, the navigation is very different from personal computers. While navigation functionality varies across mobile devices, they do not employ a mouse. Some devices use touch screens, some use thumb wheels, some use a directional key, and some use a combination of the above. Users employ a navigation technique to move the cursor on the screen, but, for many mobile devices, designers have far less control over the "tabbing" order of that movement (discussed in more detail below).

Third, mobile devices are different from computers in how the user can input information. Invariably, the size of the keyboard is one major difference, with typing commonly done with thumbs using very small buttons. Mobile device users may develop, over time, either an aversion to keyboard use or commonly-used 'shortcuts' when required to type text. If respondents are unwilling to provide text information, this may not only limit the utility of open-ended questions, but also alter closed-ended questions that offer an open-ended response option.

We investigate these differences as possible factors affecting responding in mobile web surveys. In addition, we make limited attempts to counter factors that have adverse effects on responding in mobile web surveys.

3.3.1 Small Screen Size: Failure to Obtain Necessary Information

It is possible that the very small size of the images used in the experiment reported in Section 2, led to our finding of only one difference across the seven questions. Certainly, this could not be concluded without larger sample sizes and different combinations of pictures and questions, but remains a possibility that is in need of additional experimentation.

However, the small screen also limits the amount of text information, even at small font sizes. To counter this limitation without reducing the information, such as the length of the question stem or the number of response options in a horizontal scale, the page can be allowed to extend beyond the end of the screen, requiring the respondent to scroll in order to see the entire question. While respondents can usually see that the page extends to the side, they may not be willing to expend the effort to properly perform their task.

We devised an experiment and a manipulation to explore this consequential possibility. First, respondents were randomly assigned to answer a series of eight questions on food consumption. In the first condition, they could see all seven response options (see Figure 1). Respondents in the second condition could immediately see only three response options; they could see all seven options only by "navigating" to the right (i.e., horizontally).

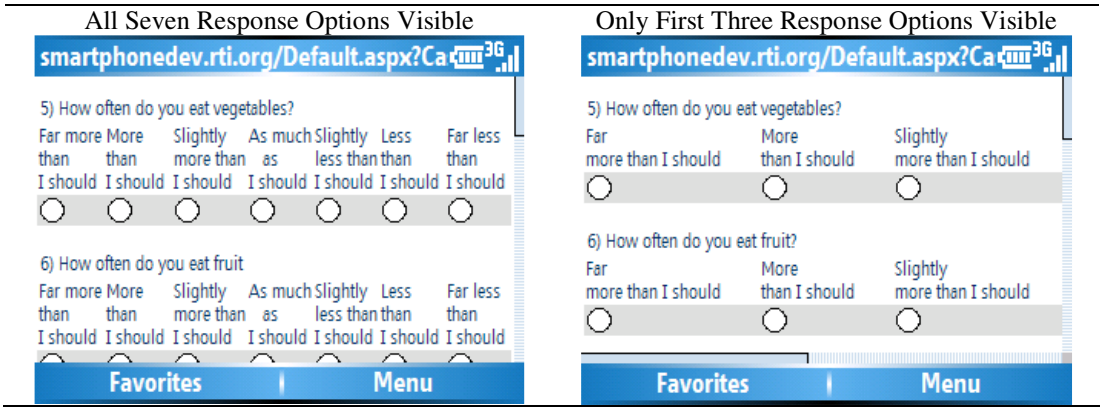

Figure 1. Screens from Experiment Varying Whether All Response Options are Visible Without Horizontal Scrolling.

No significant difference in selecting responses among the first three response options to the eight questions was found (data not shown). We also asked respondents, in a separate, subsequent question, whether they saw that the response options extended beyond the display. Among the 30 respondents assigned to the condition where only the first three response options are visible, 7 (23%) reported not noticing that there were additional response options or that it was too much effort to scroll to see them. However, 3 of these 7 actually used at least one of the nonvisible response options.

We also subjected all respondents to a question designed to lead to one answer if respondents read only the visible part, and the opposite answer if they scrolled to the right (i.e., horizontally) and read the second part of the question. Respondents were able to initially read: "Onion can increase HDL cholesterol in your body," and asked whether knowing this, they would try to eat foods with more onion, not change how much onion they eat, or try to eat foods with less onion than they have. We expected that some respondents who only read this information would respond that they would eat less onion in order to limit cholesterol. Only after scrolling to the right were they able to read: "HDL cholesterol is good for you as it removes cholesterol from arteries. (Source: www.health-cares.net)." We would expect that, given the information that onion is "good" for you, all respondents would report that they would eat more or at least the same amount of onion as they do; in fact, 3 of 61 (5%) responded that they would eat *less* onion, suggesting that they did not see, or ignored, the informative clause in the question.

3.3.2 Different Navigation: Selection of Responses in the Direct Path of the Selection Device

While requiring the respondent to scroll horizontally due to screen size is one form of navigation issue, there are other navigation features that are peculiar to mobile devices. For devices that do not have a touch-screen, moving the cursor from response option to response option might be controlled by a button, wheel, or key. Actual data input occurs by pushing a button when the cursor is "over" the desired response option, and one may have to go through all response options to reach the button for the next page. Although this may seem beneficial compared to computer administered surveys where the response options can be ignored entirely, it may present yet another challenge for mobile web survey designers: when the response options are arranged horizontally, a navigation action may move the cursor over only the first response option, depending on design—another button would be required to navigate horizontally.

Requiring more user actions has been found to lead to lower rates of using that survey feature, such as obtaining a definition with more mouse clicks (e.g., Conrad, Couper, Tourangeau and Peytchev, 2006); similarly, requiring more actions to reach response options other than the first one in horizontal scales may lead to a distortion of the response distribution towards the first response option.

We conducted an experiment with nine questions measuring mental health on a five-point scale, presented on the same page. Participants were randomly assigned to one of two versions: in the first condition, the response options were placed vertically so that moving the cursor would navigate through all response options in turn; in the second, the response options were arrayed horizontally, so that moving the cursor in the same manner moves across only the first response option to each question, requiring another button to move the cursor horizontally. The limited research on vertical versus horizontal scale orientation in computer administered web surveys suggests that there should be no difference (Thomas, Uldall and Krosnick, 2004). But, given the findings of Conrad, Couper, Tourangeau, and Peytchev (2006), our key hypothesis was that respondents in the horizontal scale condition would select the first response option at a higher rate. Similarly, we hypothesized that the means for the nine items would be lower (i.e., closer to the first option), that there would be less variability in responses, and that internal consistency would be higher.

We found no evidence for respondent unwillingness to navigate through the horizontal scale (data not shown). Twenty-three percent (31) of respondents in the vertical condition selected at least one response in the first response option, while 26% (27) in the horizontal condition did so (χ^2, p=.767). However, there was some indication that respondents may be more likely to select the first response option more than once, as none did so in the vertical condition but three selected it more than once in the horizontal condition (Mantel-Haenszel χ^2, p=.059). There were no differences in item means, no differences in variances, and no substantial difference in internal consistency (Cronbach's Alpha of .82 in the vertical versus .77 in the horizontal scale).

While there have been concerns that respondents in web panels lose motivation over time to answer thoughtfully, the incentive of the mobile device with paid voice and data plans for these participants was far higher than commonly found in web survey panels, possibly leading to higher effort in responding; this is one result that is in particular need of replication on another sample.

3.3.3 Small Keyboard Size: Avoidance of Open-ended Questions

Some survey questions require the input of numbers or text, such as open-ended questions and closed-ended questions with an "Other, please specify" response option, in addition to some procedures for logging in. Although to date there is no empirical

research comparing the length of responses to open-ended responses in mail and computer-administered web surveys, there is little reason to expect differences, other than due to design variables such as the length of the input field (Couper, Traugott and Lamias, 2001). However, keyboards on mobile devices are very small and may not even have dedicated buttons for each letter and number, likely exerting greater demand on the respondent. Making the task of text input more difficult, some respondents may not provide text information, and even select response options specifically to avoid typing. This is a source of measurement error and data quality that is rather unique to mobile devices and requires further investigation.

We designed two experiments to gauge whether respondents would select responses so as to avoid text input. In the first experiment, we asked a question on why they eat vegetables, providing two unlikely reasons (they are in the food already and they like color in their meals). We offered a third option in one of two versions: either a closed-ended "Other" or a half-open "Other, please specify:" followed by a text box. A logical expectation is that most respondents would select "Other," but when faced with having to type a response, would select one of the unlikely responses or simply leave the question unanswered.
Table 5 shows results consistent with expectations – 75% selected "Other" in the closed-ended version and only 39% selected it when it was followed by a text box, with more respondents selecting either an unlikely response or not providing a response (χ^2, $p<.05$).

Table 5 Response Distributions to: "What is the main reason you eat vegetables?"

Closed-ended (n=28)	%	Half-open (n=31)	%
"They are in my food already" or "I like color in my meals"	21%	"They are in my food already" or "I like color in my meals"	39%
"Other"	75%	"Other, please specify:" [text box]	39%
Missing	4%	Missing	23%

In another experiment, presented in Table 2, questions asked about attitudes on one valid and two fictitious issues, randomly assigning half of the participants to receive a text box next to the "Other/Don't Know" option. No significant difference was found, although on two of the questions a tendency is observed to avoid this response option when accompanied with a text field, one of the questions being the valid question. To the extent that this is a limited sample size, we find these results to be somewhat consistent with the results from the previous experiment.

4. General Discussion

Mobile web surveys can be a valuable method of data collection, but they are different in many ways and their implementation needs to be preceded by systematic research. Responses may be affected by somewhat unique features, such as the way information is displayed, the amount of information that can be presented in a screen, how the respondent navigates through the survey and provides responses, and any distractions wherever the respondent may be.

We found some evidence to support that the small screen can affect responding. When the response options extended beyond the screen, some respondents reported not having seen them. When part of the question providing strikingly different information extended beyond the screen, some respondents seemed to use only the first part of the question. Many survey questions need to provide a lot of information, just as they may require many response options, yet it may not be optimal to present such questions as they would be in other self-administered modes without careful consideration of these screen size issues.

Differences in how respondents navigate through the survey did not significantly affect responses, although there was an indication that respondents would select the same response in the path of navigation more often. Other mobile devices may induce different effects on responding through the use of other forms of navigation; indeed, some have touch screens that can be vastly different from the tabbing order tested here.

Keyboards in mobile web surveys may have a unique impact on responding to survey questions, compared to other computer-administered modes. Many survey questions ask for text information, even if it is only for one of the response options. To the degree to which the small keyboards discourage respondents from using them, necessary information may not be provided, or different (e.g., less complete, or convenient rather than sincere) responses provided. Indeed, when faced with quite implausible response options and one other alternative, the implausible response options were selected more often when the alternative required text input. Furthermore, our attempts at increasing the rate of providing open-ended responses were not successful. However, although not significant, there was some indication that respondents can be motivated to provide longer responses using the keyboard. Mobile web surveys should have much in common with their computer-administered counterparts. However, because of differences such as the size of the screen and the greater variability of situations in which the surveys are completed, mobile web surveys may not be affected by the same design features. We found very limited support for similarity in the effect of presenting images. Presenting contrasting pictures resulted in different responses to only one of seven questions. The lack of an effect of pictures can certainly be influenced by the small sample size, but may also mean that images have less of an impact on respondents when they are very small. If so, we have found an instance in which the similarities of personal computers and mobile devices do not justify the same survey designs.

Cognitive processing, as tested in these experiments, seems to be much the same in mobile web surveys as in other modes of data collection. Respondents make judgments about their response based on the frequency distribution provided in the response options, avoiding extremes on less socially desirable behaviors. They show even-handedness in responses, exhibited in different responses based on the order of questions. They also make inferences about the intent of questions, treating questions as relevant to them and providing responses to questions on fictitious issues.

There are limitations to these findings. The small sample sizes in these experiments prohibit us from full confidence in the absence of an effect when the null hypothesis is not rejected. For the same reason, respondent-related influences on usability could not be addressed, but it would be important to examine how completion of mobile web surveys is different for older and younger respondents, for example. It is possible, if not likely, that the avoidance of open-ended and half-open questions is most prevalent among the oldest adults. Similarly, the effect of such factors affecting responding in mobile web surveys may likely vary by the degree of familiarity with the technology. We had the distinct advantage of knowing that every respondent was using exactly the

same mobile device (since we provided it) and thus having to make same navigation choices and actions; in any study but a small and tightly-controlled pilot such as ours, there will be great variation across devices.

Further research is needed into the identification of factors affecting response processes in mobile web surveys, as well as into the development of theory-guided methods to alleviate undesirable effects. While such research is imperative, conducting it sooner may not only improve early studies using this mode, but also preserve respondent faith in the intentions of mobile web surveys.

Acknowledgements

The authors thank the smartphone panel team—LaChanda Carter, Derek Coombs, Joe Eyerman, Jay Levinsohn, Joseph McMichael, Brian Olive, Jamie Ridenhour, David Roe, and Ann Zhang.

References

Brick, J. M., P. D. Brick, S. Dipko, S. Presser, C. Tucker and Y. Yuan (2007). "Cell Phone Survey Feasibility in the U.S.: Sampling and Calling Cell Numbers Versus Landline Numbers." Public Opinion Quarterly **71**(1): 23-39.

Chaiken, S. (1987). The Heuristic Model of Persuasion. Social Influence: The Ontario Symposium. M. P. Zanna, J. M. Olson and C. P. Herman. Mahwah, NJ, Lawrence Erlbaum Associates. **5:** 3-39.

Conrad, F. G., M. P. Couper, R. Tourangeau and A. Peytchev (2006). "Use and Non-Use of Clarification Features in Web Surveys." Journal of Official Statistics **22**(2): 245-269.

Couper, M. P., F. G. Conrad and R. Tourangeau (2007). "Visual Context Effects in Web Surveys." Public Opinion Quarterly **71**(4): 623-634.

Couper, M. P., R. Tourangeau, F. Conrad and S. Crawford (2004). "What They See Is What We Get: Response Options for Web Surveys." Social Science Computer Review **22**(1): 111-127.

Couper, M. P., M. W. Traugott and M. J. Lamias (2001). "Web Survey Design and Administration." Public Opinion Quarterly **65**(2): 230-253.

DeRouvray, C. and M. P. Couper (2002). "Designing a Strategy for Reducing "No Opinion" Responses in Web-Based Surveys." Social Science Computer Review **20**(1): 3-9.

Dillman, D. A. (2000). Mail and Internet Surveys: The Tailored Design Method. New York, Wiley.

Fuchs, M. (2001). Screen Design in a Web Survey. Proceedings of Paper presented at the annual meeting of the American Association for Public Opinion Research, Montréal, Quebec.

Grice, H. P. (1975). Logic and Conversation. Syntax and Semantics. P. Cole and J. L. Morgan. New York, Academic Press. **3:** 41-58.

Hardwick, M. E., P. A. Pulido and W. S. Adelson (2007). "The Use of Handheld Technology in Nursing Research and Practice." Orthopaedic Nursing **26**(4): 251-255.

Hyman, H. H. and P. B. Sheatsley (1950). The Current Status of American Public Opinion. The Teaching of Contemporary Affairs: Twenty-First Yearbook of the National Council for the Social Studies. J. C. Payne. Washington, D.C., National Council for the Social Studies.

Jones, M., G. Marsden, N. Mohd-Nasir, K. Boone and G. Buchanan (1999). "Improving Web Interaction on Small Displays." Computer Networks-the International Journal of Computer and Telecommunications Networking **31**(11-16): 1129-1137.

Mooney, G. M., B. Rogers and D. Trunzo (2003). Examining the Effect of Error Prompting on Item Nonresponse and Survey Nonresponse in Web Surveys. Proceedings of American Association for Public Opinion Research, 58th Annual Conference, Nashville, Tennessee.

Nielsen Mobile (2008). Critical Mass: The Worldwide State of the Mobile Web. **2008**.

Okazaki, S. (2007). "Assessing Mobile-Based Online Surveys: Methodological Considerations and Pilot Study in an Advertising Context." International Journal of Market Research **49**(5): 651-674.

Parush, A. and N. Yuviler-Gavish (2004). "Web Navigation Structures in Cellular Phones: The Depth/Breadth Trade-Off Issue." International Journal of Human-Computer Studies **60**(5-6): 753-770.

Petty, R. E. and J. T. Cacioppo (1986). Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York, Springer-Verlag.

Peytchev, A. A. (2007). Participation Decisions and Measurement Error in Web Surveys. Survey Methodology. Ann Arbor, University of Michigan. **Ph.D.**

Schwarz, N., H. J. Hippler, B. Deutsch and F. Strack (1985). "Response Scales - Effects of Category Range on Reported Behavior and Comparative Judgments." Public Opinion Quarterly **49**(3): 388-395.

Sudman, S., N. M. Bradburn and N. Schwarz (1996). Thinking About Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco, Jossey-Bass Publishers.

Sweeney, S. and F. Crestani (2006). "Effective Search Results Summary Size and Device Screen Size: Is There a Relationship?" Information Processing & Management **42**(4): 1056-1074.

Thomas, R. K., B. Uldall and J. A. Krosnick (2004 of Conference). How Many Are Too Many? Number of Response Categories and Validity. Paper Presented at American Association for Public Opinion Research, Phoenix, AZ.

Tourangeau, R., M. P. Couper and F. Conrad (2004). "Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions." Public Opinion Quarterly **68**(3): 368-393.

van Schaik, P. and J. Ling (2007). "Design Parameters of Rating Scales for Web Sites." ACM Transactions on Computer-Human Interactions **14**(1): Article 4.

Weber, B. A., H. Yarandi, M. A. Rowe and J. P. Weber (2005). "A Comparison Study: Paper-Based Versus Web-Based Data Collection and Management." Applied Nursing Research **18**(3): 182-185.