Gridlocked: The impact of adapting survey grids for smartphones

Ashley Richards*, Rebecca Powell, Joe Murphy, Mai Nguyen, Shengchao Yu

Keywords: data quality, responsive design, web survey, smartphone, mobile, grid

DOI: [10.29115/SP-2016-0016](https://doi.org/10.29115/SP-2016-0016)

Survey Practice

Vol. 9, Issue 3, 2016

Paper and web surveys often include grid-style questions formatted to save space and avoid repetition. However, this format is often discouraged by methodologists because of data quality concerns, particularly when respondents are using small screens (e.g., smartphones). In the World Trade Center Health Registry's Wave 4 survey, we used grids to maintain comparability with prior waves. However, due to the rising number of respondents using smartphones to complete web surveys, we used responsive web design programming to automatically reformat grids into a series of individual items when a small screen device was detected. This method allowed us to retain grids used in previous years, yet address the issue of grids displaying poorly on mobile devices. We compared indicators of data quality (e.g., missing data, straight-lining) across grid formats to see whether the smartphone-optimized version suggests poorer, equal, or better data quality than the traditional grid. We also compared consistency with data collected in previous waves of the survey. We found some evidence that the optimized grid format improved data quality, and the benefits we observed may even suggest that some variant of the mobile-optimized format should be considered for all devices, regardless of screen size.

INTRODUCTION

Grid questions are often used to save space in surveys. While the decision to use grids may be justified in certain situations, the survey literature includes a number of examples of negative data quality outcomes resulting from the use of grids (e.g., Toepoel et al. 2009; Tourangeau et al. 2004). This is especially true in recent years, as the percentage of surveys taken on smartphones is rising (Link et al. 2014), and grids display poorly on small screens. These types of challenges when presenting grids on smartphones have left researchers looking for alternate solutions. In this paper, we introduce a "stacked" format that presents each row of a grid separately for smartphone screens. To determine if people respond differently to the small screen's stacked grid compared to the large screen's nonstacked (i.e., traditional) grid, we compare outcomes and indicators of data quality between the two versions.

BACKGROUND AND LITERATURE

A benefit of grids is they save space by avoiding repetition. However, the

* Institution: RTI International

literature suggests grids can have a number of drawbacks. Completing a grid is cognitively challenging (Dillman, Smyth, and Christian 2014), and the task is even more complicated when horizontal or vertical scrolling is required to view the entire grid. An additional drawback of grids is that they tend to result in faster survey completion times (e.g., Couper et al. 2013), which may increase measurement error (Peytchev 2005).

However, many of the findings on grids are mixed. Some studies find higher inter-item correlations (straight-lining) in grids (e.g., Toepoel et al. 2009), while others find no difference (e.g., Callegaro et al. 2009). Furthermore, some studies suggest rates of missing data may be higher in grids (e.g., Toepoel et al. 2009), while others find the opposite (Couper, Traugott, and Lamais 2001).

Dillman and colleagues (2014) suggest presenting grid items as a series of individual questions when mobile responses are anticipated. This is especially critical because of continuously rising rates of smartphone and tablet use – currently 68 percent of U.S. adults have a smartphone and 45 percent have a tablet (Pew Research Center 2015). This increase correlates with a rise in survey participation via mobile devices (Link et al. 2014). Nearly every web survey today includes mobile respondents, and the design of surveys must acknowledge and accommodate users of these devices.

Despite recommendations to avoid grids, sometimes their use is justified. It may be important to retain grids in longitudinal studies that have used grids previously or in mail surveys with page constraints. For this web and mail survey, we attempted to optimize grid display on mobile devices, so the decision to use grids would not automatically sacrifice comparability across devices (e.g., smartphone vs. desktop) or modes (i.e., web vs. mail).

METHODS

In this paper, we compare traditional and mobile-optimized grid formats on the World Trade Center (WTC) Health Registry's 2015 health survey. The WTC Health Registry was established in 2002 to monitor the long-term physical and mental health of people exposed to the September 11, 2001 terrorist attack on the WTC in New York City, and to better assess the post-disaster health care needs of survivors (Brackbill et al. 2009; Farfel et al. 2008).

Four waves of the survey have taken place since the WTC Health Registry was created in 2002. The fourth wave was launched in early April 2015. It offered two modes: web and mail. This paper presents findings based on 14,613 web responses that were received through August 2015 before nonresponse follow-up efforts began.

This survey contained five grids¹ to assess: (1) health condition diagnoses, (2)

¹ In the Appendix, we present exact question wording and visual structure of the questions in Figures 8–12.

post-traumatic stress disorder (PTSD) symptoms, (3) depression symptoms, (4) psychological distress symptoms, and (5) availability of social support. We used grids in the survey to maintain the original format of validated scales. Furthermore, grids were the most practical format – particularly in the mail mode, which had page constraints – and we wanted to ensure comparability across modes of the survey, as well as to previous waves of the survey, which used grids. We anticipated that this wave would have a greater number of mobile respondents than earlier waves, so we optimized the display of grids to account for the respondent’s screen size.

The survey was programmed to modify the display of grid questions based on the respondent’s screen width. For devices whose width was greater than 760 pixels (the approximate width of a phone in landscape mode), grid questions were rendered in a traditional grid format (Figure 1). For smaller-width devices (smartphones or small tablets), grids were rendered in a “stacked format” that presented each row separately (Figure 2). The stacked format improves the display on small screens by eliminating the need for horizontal scrolling. We investigated if and how data quality differed between the traditional and stacked versions of the grids by examining three measures of data quality: (1) mean responses across the items, (2) missing data, and (3) straight-lining.

NYC NEW YORK CITY DEPARTMENT OF HEALTH and MENTAL HYGIENE

World Trade Center Health Registry
2015 Health Survey

Help Log Out

22% Complete

Q30. Have you ever been told by a doctor or other health professional that you had any of these conditions? Note: Cancer is covered later in this survey.

	Yes	No
Hypertension, or high blood pressure	☐	☐
High cholesterol	☐	☐
Angina, or angina pectoris	☐	☐
Heart attack, or myocardial infarction	☐	☐
Coronary heart disease	☐	☐
Stroke	☐	☐
Diabetes, or sugar diabetes	☐	☐
Asthma	☐	☐
Chronic bronchitis	☐	☐
Emphysema, or COPD	☐	☐
Reactive airways dysfunction syndrome, or RADS	☐	☐
Sarcoidosis	☐	☐
Pulmonary fibrosis	☐	☐
Asbestosis	☐	☐

Figure 1 Grid format viewed on a computer.

NYC NEW YORK CITY DEPARTMENT OF HEALTH AND MENTAL HYGIENE

World Trade Center Health Registry
2015 Health Survey

Help Log Out

Q30. Have you ever been told by a doctor or other health professional that you had any of these conditions? Note: Cancer is covered later in this survey.

Hypertension, or high blood pressure

Yes	No
☐	☐
Clear Answer	

High cholesterol

Yes	No
☐	☐
Clear Answer	

Angina, or angina pectoris

Yes	No
☐	☐

Figure 2 Stacked format viewed on a smartphone.

ANALYSIS

As we could not randomly assign respondents to a device, we needed to account for the respondent differences due to self-selection in the analysis. Although there are multiple methods to accomplish this, we chose to control for respondents' characteristics in regression models as it is a well-established, robust statistical method. Specifically, we analyze the data through predicted probabilities from logistic and ordinary least squares regression models.

Since this survey is longitudinal, we can also account for the lack of experimental design by comparing the results in the current survey (Wave 4 – mobile optimized) to the results in the previous wave. We compared across waves responses to the health conditions grid, because this was the only grid with a lifetime reference period.

RESULTS

DEMOGRAPHICS

Out of the 14,613 web respondents, 11 percent completed the survey on a small device such as a smartphone. Smartphone respondents were statistically younger and less educated than larger screen respondents. Due to these demographic differences, we control for age, education, and gender when analyzing the differences in responses across the two grid formats.

RESPONSE DIFFERENCES

We began by comparing response distributions for grid items across formats (traditional vs. stacked). These grids were all ordered from positive to negative,

with fewer symptoms on the left and more on the right.² For four of the grids, respondents who viewed the stacked format selected on average more “negative” responses than those viewing the traditional format, meaning they reported having fewer sources of social support and more symptoms of PTSD, depression, and psychological distress. For every item in these grids, the response differences across formats followed the same pattern of the stacked format resulting in more negative selections. Figure 3 displays this pattern for the PTSD grid; the depression, psychological distress, and social support grids demonstrate the same pattern and are available upon request. There was no clear pattern in the remaining grid, which asks about health conditions with which the respondent has ever been diagnosed (see Figure 4).

Figure 3 PTSD grid: average response selected for each item by grid format.

² From left to right, response options were ordered from positive to negative and ranged from “not at all” (score of 1) to “extremely,” (score of 5) or “none of the time” (score of 1) to “all of the time” (score of 5). The social support grid was opposite in that from left to right its order was negative to positive, i.e., “not at all” to “nearly every day.” To maintain consistency across the grids, the social support scale was reversed in the analysis so that “nearly every day” was a score of 1 and “not at all” was a score of 5.

Figure 4 Health conditions grid: average response selected for each item by grid format.

DATA QUALITY

We examined rates of missing data and straight-lining to get an indication of data quality in the traditional vs. stacked formats. In all but the social support grid, the predicted probability that respondents would leave at least one item unanswered is significantly higher in the traditional grid compared to the stacked grid. These differences were statistically significant. Figure 5 compares the predicted probabilities of respondents missing an item for each grid by grid format.

Figure 5 Predicted probability of respondents skipping at least one item by grid format.

The predicted probability that respondents would skip an entire grid followed the same pattern: skipping an entire grid was more likely among those viewing the traditional grid, and this was true for all grids except social support (see Figure 6). Skipping an entire grid was rare, with predicted probabilities ranging from 0.0025 to only 0.0047.

Figure 6 Predicted probability of skipping the entire grid by grid format.

Next, we looked at straight-lining to see if one of the grid formats was more likely to result in respondents selecting the same response option for all items in a grid. In all five grids, respondents were significantly more likely to straight-line when presented with the traditional format compared to the stacked format ($p < 0.05$; see Figure 7).

Figure 7 Predicted probability of straight-lining by grid format.

DIFFERENCES OVER TIME

Finally, we compared the health conditions reported during the current and previous waves to determine if the stacked format affected accuracy. We gauged accuracy as the absence of discrepancies across the two waves, using only conditions reported as having a diagnosis year of 2010 or earlier, as these conditions should have been reported in both waves. Overall, respondents had an average of 1.6 discrepancies, meaning they reported these conditions in only one wave instead of in both waves. However, the number of discrepancies per respondent was not statistically different across the two grid formats.

SUMMARY AND DISCUSSION

Four of the five grids had a clear trend of more negative responses being selected in the stacked format. These grids were similar in structure (4–5 response options) and topic (psychological symptoms). It is not surprising that these grids showed a response pattern that the health conditions grid did not, because these grids are scales and we would expect respondents to select similar responses across items within each grid.

For the health conditions grid, the probability of respondents skipping items in the stacked format was less than half of that for the traditional grid format. We speculate the following may have contributed to the higher rate of item missingness in the traditional format:

1. At 22 items, the grid's length may be perceived as intimidating and burdensome, particularly when all items are visible as they are on a

large screen (traditional format).

2. The stacked format breaks out items so they are viewed individually, making it less likely respondents will skip an item (i.e., row in the grid) without realizing it.
3. Some respondents may have treated the forced choice yes/no format in the health conditions grid as though it were a checkbox format (i.e., they may have left items blank rather than selecting “No”).

The likelihood of skipping an entire grid by format varied across grids. Respondents who viewed the stacked format were significantly more likely to skip the entire social support grid, while those viewing the traditional grid were significantly more likely to skip the other grids. The social support grid appears late in the survey, so fatigue among mobile respondents may have driven the difference between the traditional and stacked versions of this grid.

We found that straight-lining was significantly more likely across all grids in the traditional format than the stacked format. This suggests that straight-lining can be reduced by forcing respondents to consider each item individually, as is done in the stacked format. Although these data quality measures are statistically significant, it is important to note that the impact of some measures (e.g., straight-lining) may be greater than the impact of others with lower predicted probabilities (e.g., skipping an entire grid).

We found that the inconsistencies between the current and previous waves were roughly equally common in both the traditional and stacked grid formats. One format did not appear to elicit more consistent responses.

A limitation of our approach is that we were unable to randomly assign sample members to the stacked (mobile) and traditional (nonmobile) versions of the grids. As a result, we do not know to what extent observed differences are due to layout or device vs. characteristics of respondents using each type of device.

We know that grids have both advantages (e.g., space efficiency, faster completion time) and disadvantages (e.g., straight-lining, missing data, cognitive challenge). As this was the fourth wave of the survey, we determined the need to make longitudinal comparisons to earlier grids outweighed the disadvantages of using grids. However, in earlier waves of the survey, smartphones were much less prevalent and we assume mobile responding was uncommon, so the issue of grids displaying poorly on smartphones was less of a concern.

Given that there will be respondents who respond via smartphone, researchers should use a grid format that works on small screens, while maintaining compatibility with the traditional grid to reduce mode effects. We recommend using the stacked format, as it (1) presents items in a grid-like fashion but formats them for optimal mobile usability and (2) has no consistent evidence

of decreased data quality, and in fact, we found evidence of slightly increased data quality in some circumstances. Data quality is an important consideration above and beyond space efficiency and survey administration time. The benefits we observed with the stacked version of the grid may even suggest that some variant of the stacked format should be considered for all devices, regardless of screen size.

APPENDIX

Have you ever been told by a doctor or other health professional that you had any of these conditions? If yes, please provide the year you were first told you had that condition.

	No	Yes		Year first told
a. Hypertension, or high blood pressure	☐	☐	→	_____
b. High cholesterol	☐	☐	→	_____
c. Angina, or angina pectoris	☐	☐	→	_____
d. Heart attack, or myocardial infarction	☐	☐	→	_____
e. Coronary heart disease	☐	☐	→	_____
f. Stroke	☐	☐	→	_____
g. Diabetes, or sugar diabetes	☐	☐	→	_____
h. Asthma	☐	☐	→	_____
i. Chronic bronchitis	☐	☐	→	_____
j. Emphysema, or COPD	☐	☐	→	_____
k. Reactive airways dysfunction syndrome, or RADS	☐	☐	→	_____
l. Sarcoidosis	☐	☐	→	_____
m. Pulmonary fibrosis	☐	☐	→	_____
n. Asbestosis	☐	☐	→	_____
o. Sleep apnea, or obstructive sleep apnea	☐	☐	→	_____
p. Gastroesophageal reflux disease, or GERD	☐	☐	→	_____
q. Thyroid disease	☐	☐	→	_____
r. Peripheral neuropathy	☐	☐	→	_____
s. Rheumatoid arthritis	☐	☐	→	_____
t. Other auto-immune disorders (e.g., lupus, MS, ALS, scleroderma, or polymyositis)	☐	☐	→	_____
u. *Other disease (excluding cancer), Please specify: _____	☐	☐	→	_____
v. *Other disease (excluding cancer), Please specify: _____	☐	☐	→	_____

Figure 8 Health conditions grid.

Over the last 2 weeks, how often have you been bothered by any of the following problems?

	Not at all	Several days	More than half the days	Nearly every day
a. Little interest or pleasure in doing things	☐	☐	☐	☐
b. Feeling down, depressed, or hopeless	☐	☐	☐	☐
c. Trouble falling or staying asleep, or sleeping too much	☐	☐	☐	☐
d. Feeling tired or having little energy	☐	☐	☐	☐
e. Poor appetite or overeating	☐	☐	☐	☐
f. Feeling bad about yourself, or that you are a failure or have let yourself or your family down	☐	☐	☐	☐
g. Trouble concentrating on things, such as reading the newspaper or watching television	☐	☐	☐	☐
h. Moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual	☐	☐	☐	☐

Figure 9 PTSD grid.

In the <u>last 30 days</u> , how much have you been bothered by the following problems?					
	Not at all	A little bit	Moderately	Quite a bit	Extremely
a. Repeated, disturbing memories, thoughts, or images of the events of 9/11	☐	☐	☐	☐	☐
b. Repeated, disturbing dreams of the events of 9/11	☐	☐	☐	☐	☐
c. Suddenly acting or feeling as if the events of 9/11 were happening again (as if you were reliving it)	☐	☐	☐	☐	☐
d. Feeling very upset when something reminded you of the events of 9/11	☐	☐	☐	☐	☐
e. Having physical reactions (e.g., heart pounding, trouble breathing, sweating) when something reminded you of the events of 9/11	☐	☐	☐	☐	☐
f. Avoiding thinking about or talking about the events of 9/11 or avoiding having feelings related to it	☐	☐	☐	☐	☐
g. Avoiding activities or situations because they remind you of the events of 9/11	☐	☐	☐	☐	☐
h. Trouble remembering the important parts of the events of 9/11	☐	☐	☐	☐	☐
i. Loss of interest in activities that you used to enjoy	☐	☐	☐	☐	☐
j. Feeling distant or cut off from other people	☐	☐	☐	☐	☐
k. Feeling emotionally numb or being unable to have loving feelings for those close to you	☐	☐	☐	☐	☐
l. Feeling as if your future will somehow be cut short	☐	☐	☐	☐	☐
m. Trouble falling or staying asleep	☐	☐	☐	☐	☐
n. Feeling irritable or having angry outbursts	☐	☐	☐	☐	☐
o. Having difficulty concentrating	☐	☐	☐	☐	☐
p. Being "super alert" or watchful or on guard	☐	☐	☐	☐	☐
q. Feeling jumpy or easily startled	☐	☐	☐	☐	☐

Figure 10 Depression grid.

During the <u>last 30 days</u> , about how often did you feel:					
	None of the time	A little of the time	Some of the time	Most of the time	All of the time
a. So sad that nothing could cheer you up?	☐	☐	☐	☐	☐
b. Nervous?	☐	☐	☐	☐	☐
c. Restless or fidgety?	☐	☐	☐	☐	☐
d. Hopeless?	☐	☐	☐	☐	☐
e. That everything was an effort?	☐	☐	☐	☐	☐
f. Worthless?	☐	☐	☐	☐	☐

Figure 11 Psychological distress grid.

How <u>often</u> is someone available:					
	None of the time	A little of the time	Some of the time	Most of the time	All of the time
a. To take you to the doctor if you need to go?	☐	☐	☐	☐	☐
b. To have a good time with?	☐	☐	☐	☐	☐
c. To hug you?	☐	☐	☐	☐	☐
d. To prepare your meals if you are unable to do it yourself?	☐	☐	☐	☐	☐
e. To understand your problems?	☐	☐	☐	☐	☐

Figure 12 Social support grid.

REFERENCES

- Brackbill, R.M., J.L. Hadler, L. DiGrande, C.C. Ekenga, M.R. Farfel, S. Friedman, S.E. Perlman, et al. 2009. "Asthma and Posttraumatic Stress Symptoms 5 to 6 Years Following Exposure to the World Trade Center Terrorist Attack." *The Journal of the American Medical Association* 302 (5): 502–16. <https://doi.org/10.1001/jama.2009.1121>.
- Callegaro, M., J. Shand-Lubbers, and J.M. Dennis. 2009. "Presentation of a Single Item versus a Grid: Effects on the Vitality and Mental Health Subscales of the Sf-36v2 Health Surveys." Presented at the Annual Meeting of the American Association for Public Opinion Research. Hollywood, FL.
- Couper, M.P., R. Tourangeau, F.G. Conrad, and C. Zhang. 2013. "The Design of Grids in Web Surveys." *Social Science Computer Review* 31 (3): 322–45.
- Couper, M.P., M.W. Traugott, and M.J. Lamais. 2001. "Web Survey Design and Administration." *Public Opinion Quarterly* 65 (2): 230–53.
- Dillman, D.A., J.D. Smyth, and L.M. Christian. 2014. *Internet, Phone, Mail, and m-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons.
- Farfel, M., L. DiGrande, R. Brackbill, A. Prann, J. Cone, S. Friedman, D.J. Walker, et al. 2008. "An Overview of 9/11 Experiences and Respiratory and Mental Health Conditions among World Trade Center Health Registry Enrollees." *Journal of Urban Health* 85 (6): 880–909. <https://doi.org/10.1007/s11524-008-9317-4>.
- Link, M.W., J. Murphy, M.F. Schober, T.D. Buskirk, J.H. Childs, and C.L. Tesfaye. 2014. "Mobile Technologies for Conducting, Augmenting, and Potentially Replacing Surveys: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research." http://www.aapor.org/AAPOR_Main/media/MainSiteFiles/REVISED_Mobile_Technology_Report_Final_revised10June14.pdf.
- Pew Research Center. 2015. "Technology Device Ownership." <http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/>.
- Peytchev, A. 2005. "How Questionnaire Layout Induces Measurement Error." Presented at the Annual Meeting of the American Association for Public Opinion Research. Miami Beach, FL.
- Toepoel, V., M. Das, and A. Van Soest. 2009. "Design of Web Questionnaires: The Effects of the Number of Items per Screen." *Field Methods* 21 (2): 200–213.
- Tourangeau, R., M.P. Couper, and F.G. Conrad. 2004. "Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions." *Public Opinion Quarterly* 68 (3): 368–93.