Improving the Mobile Survey Experience:
A closer look at scale orientation, grids and modality effects

by Keith Chrzan, Chief Research Officer, Maritz Research and
Ted Saunders, Online Research Manager, Maritz Research
The increase in the ownership and usage of smartphones presents a paradigm shift in how we view our web interactions with survey participants. Not surprisingly, most studies using email-invitations have a subset of respondents who attempt to take surveys using their mobile phone’s web browser.

Form factors and technology related to mobile phones can make survey participation challenging. Maritz Research recently completed a research on research study addressing this topic. The goal was to identify the best way to present scales (horizontal versus vertical) and grids (standard multi-attribute grids versus grids broken into multiple scaled questions with one attribute per page). Outcomes were measured by respondent engagement, satisfaction and data quality. The study also analyzed score differences and demographic differences between PC and mobile web respondents, enabling us to determine whether the bias in ratings we have seen in other studies (taken on both mobile phones and PCs) is attributable to true mode differences or possibly interpersonal differences between the types of respondents who choose to use one device versus another. Finally, we analyzed whether data collected on mobile phones, where the possibility of distraction seems greater, is worse than data collected on PCs.

The web-based survey was administered to an online panel comprised of 1,506 recent (past 90 days) video services users (e.g. Netflix, Redbox, Cable pay-per-view). As an additional qualifier, these respondents previously expressed a willingness to complete surveys using web interfaces on their mobile phones. Nearly 60% (901) completed the survey using their mobile phones, while 605 completed the survey on a PC. The initial invitation did not specifically instruct respondents to complete the survey from any particular type of device for fear that such instruction would impact respondent’s expectations of the survey.

For each respondent, the interview software randomly selected one of the services the respondent had used. A respondent then rated that video service using six different rating scale measures of customer satisfaction and loyalty. About 67% (1,004) respondents received the satisfaction and loyalty rating scales in a horizontal format and 502 received them in a vertical format. Respondents also rated the performance of their selected brands on nine attributes, of which a third of respondents received the attributes in a typical horizontal grid format, a third received them as separate questions in a horizontal format and a third received them as separate questions in a vertical format. Respondents also provided a behavioral prediction and a rating of brand liking for a total of 17 substantive measurements. Finally, at the end of the survey, respondents rated their satisfaction with the survey on a standard five-point, fully anchored rating scale.

Cell definitions and sample size by experimental treatment appear as Exhibit 1. Screen shots of the different modes and formats appear in the Appendix.

In order to measure test-retest reliability, we re-interviewed 1,187 (79%) of the respondents seven days after they completed the initial survey and asked them the four satisfaction questions. The re-contact invitation identified the type of device (PC or mobile phone) the
respondent used to complete the initial interview and requested they use the same device to complete the re-contact survey.

This research design allows us to assess:

- Respondents’ mode preference and satisfaction
- The actual and perceived survey length and drop-off rates by mobile-versus-PC and by grid-versus-horizontal-versus-vertical
- The test/re-test reliability by mobile-PC and by horizontal-vertical
- Scale use differences by mobile-PC and by horizontal-vertical-grid
- Incidence of straightlining by mobile-PC and by horizontal-vertical-grid

The results of our various tests are as follows:

**Respondent Characteristics**

Over two thirds (68%) of respondents were female. One in 10 were age 25 or younger; 45% were 25-44 years olds and another 45% were over 44. Most respondents (71%) reported being employed full time and the sample reported a median household income of $80,000. A majority of respondents (57%) reported having at least a college degree. Between 10% and 15% of respondents rated each of five brands (Blockbuster, Netflix DVD, Netflix streaming, Cable pay-per-view and Redbox) while between 5% and 10% of respondents rated each of the five smaller brands (satellite pay-per-view, iTunes rental, Hulu Plus, Amazon Instant rental and Crackle.com).

Analysis of variance showed substantive survey responses related significantly to respondent sex and education level and to the brand being rated. Below we report results based on raw survey data but we also ran our statistical tests two ways to confirm their robustness to sample differences. We ran the tests using data rim weighted to comparability on these three variables and we ran them as part of multifactorial analysis of variance. Both the weighting and the statistical adjustments produced the same significance test results as the raw data.

**Respondent mode preference**

Of the 901 mobile web respondents to the initial survey, 484 took the re-contact survey on their mobile phones (187 switched to PC and do not qualify for this analysis). Of the 605 PC respondents, 492 took the re-contact survey on a PC and 24 on a mobile phone. Note we had a total re-contact rate of 86% for initial PC respondents and only 75% for initial mobile respondents. Also, only 72% of the initial mobile respondents who opted to take the re-contact did so using their mobile phones while 95% of initial PC respondents taking the re-contact survey did so using their PCs. The significantly greater re-contact rate of initial PC respondents and the significantly greater mode switching behavior among the initial mobile respondents suggests the mobile web survey may have been a less pleasant experience for respondents.
Completion Rates
As we have experienced on most email based studies, completion percentages* from surveys started on mobile devices were lower than surveys started on PCs. This was true for all three versions of the survey as evidenced in the table below.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Horizontal Grid</th>
<th>Horizontal Separate</th>
<th>Vertical Separate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC Web</td>
<td>81.9%</td>
<td>81.1%</td>
<td>79.2%</td>
</tr>
<tr>
<td>Mobile Web</td>
<td>76.5%</td>
<td>74.1%</td>
<td>74.4%</td>
</tr>
</tbody>
</table>

*(as measured by those respondents who clicked on the URL to begin the survey and progressed through the survey to complete the final closed-ended question)

We analyzed completion percentage and survey length by question and found the majority of the variance in completion percentage between PC and mobile occurred in the first two questions of the survey. The completion time over the first two questions was 28 seconds on mobile devices compared to 18 seconds on PCs, as the less sophisticated mobile browsers took longer to load. The question complete times over the remainder of the survey continued to be longer on mobile devices but the completion rates/drop off rates did not show significant differences between mobile and PC respondents. We had expected to see higher drop off rates from the mobile respondents on the grid question but that did not happen. While the grid took significantly longer to complete on a mobile device, the drop off rates were comparable among the mobile and PC respondents. It appears that mobile abandons are largely tied to survey load times rather than complex question formats and that if respondents dislike the mobile survey experience due to the load time, they are likely to drop off early in the survey.

In this study, we detected the type of device respondents were using upon entry to the survey and we routed respondents on mobile devices to versions of the survey optimized to their phone (separate survey renderings for Android, Blackberry and iPhone users). Had we not employed ‘mobile friendly’ versions of the survey we expect a larger difference in the completion percentages between mobile and PC respondents would have resulted.

Respondent Satisfaction
Respondents’ satisfaction with the survey did not depend on whether they completed it on a PC or on mobile web browser. Nor did whether respondents saw horizontally or vertically oriented scales or whether their battery of attributes appeared in a grid or as separate questions. Respondents in all experimental conditions reported satisfaction averaging between 4.51 and 4.59.

Survey length
We have both actual and respondent-perceived/reported measures of survey duration. Because respondents can pause taking the survey and return to complete it later, we compare median rather than mean survey length. Surveys are significantly shorter for PC respondents than for mobile web, and for respondents receiving their attribute battery as a grid rather
than broken out into separate horizontal or vertical scales. Respondents’ perceptions show the same relative differences (though, perhaps because the invitation promised a five minute survey, they thought the survey took over five minutes):

<table>
<thead>
<tr>
<th></th>
<th>Actual Minutes (mean)</th>
<th>Perceived Minutes (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC web</td>
<td>2.60</td>
<td>5.31</td>
</tr>
<tr>
<td>Mobile web</td>
<td>3.77</td>
<td>5.74</td>
</tr>
<tr>
<td>Grid</td>
<td>2.85</td>
<td>5.18</td>
</tr>
<tr>
<td>Horizontal</td>
<td>3.30</td>
<td>5.59</td>
</tr>
<tr>
<td>Vertical</td>
<td>3.53</td>
<td>5.93</td>
</tr>
</tbody>
</table>

With respect to survey length, our hypotheses about vertical presentation of scales and about breaking apart grids (at least shorter ones like those with nine attributes) appear to be false.

**Data Quality Issues**

When a new mode emerges, researchers naturally wonder about the quality of the data that mode can produce. Many researchers were skeptical about replacing mail surveys with phone and then both with web-based surveys. Now, one might wonder whether we can trust mobile-web data more or less than PC web data. This research allows us to investigate several aspects of data quality.

**Bias in rating scales**

As noted, Maritz has consistently seen mobile web respondents give less favorable overall satisfaction ratings than PC web respondents. Weighing by a handful of demographics (age, sex, income) has reduced but not eliminated this bias, leading us to wonder whether the residual bias owes to modality or to unmeasured demographic or psychographic differences. In the current survey, all respondents had professed willingness to answer surveys via their mobile phones, largely ruling out unmeasured demographic and psychographic differences as an explanation. If we see significant differences in how respondents use rating scales we can attribute them to a difference caused by modality itself. If not, the differences we see in other studies may owe to unobserved demographic or psychographic differences, not to modality. However, this conclusion merits more testing.

Our evidence favors the latter. For a total of 17 substantive questions (16 of them rating scale measures) including eight overall measures and nine attributes, analysis of variance shows NO significant differences depending on whether a respondent answers via her PC or her mobile phone. We do, however, see significant differences on four of the 17 measures depending on whether they use a vertical, horizontal or grid format where scores in the horizontal format are significantly higher than for vertical or grid format.
Test/re-test reliability
A basic measure of data quality is the ability to measure the same thing at different times and get the same answer. Psychometricians call this “test/re-test reliability.” In this study we collected an overall satisfaction measure in the initial survey and again seven days later in the re-contact survey. The table below shows correlations between the initial survey data and the re-contact data for our overall satisfaction question:

<table>
<thead>
<tr>
<th>Mobile reliability</th>
<th>PC reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>.73</td>
<td>.66</td>
</tr>
</tbody>
</table>

Far from giving us data to worry about, the quality of mobile web data appears quite good. We can also assess the reliability of satisfaction ratings collected using horizontal and vertical layout of the scales. For the satisfaction question, the horizontal scale provides a significantly more reliable measure than the vertical scale:

<table>
<thead>
<tr>
<th>Vertical</th>
<th>Horizontal</th>
</tr>
</thead>
<tbody>
<tr>
<td>.65</td>
<td>.73</td>
</tr>
</tbody>
</table>

Straightlining
The study also measured the extent to which respondents straightline their attribute questions. No significant difference straightlining among respondents who answer by PC or by mobile phone respondents receive horizontal or vertical scale presentation, or to whether the attributes appear as a grid or as separate questions.

Discussion
We sampled from a subset of panelists who had identified themselves as likely to complete surveys on their mobile web browsers in an effort to minimize self-selection effects. We did allow respondents to decide how to take the survey, however, so we did not completely eliminate mode self-selection. We further ensured the comparability of the groups in two ways: by running statistical comparisons weighted by significant demographic and brand usage differences and by running multi-factorial analysis of variance. Unlike every previous study we have analyzed, we found no differences in substantive rating scale responses owing to modality (PC versus mobile web). This may indicate the differences we usually see owed to respondent differences and not specifically to the mechanics of different survey modes; if so, this result bolsters the case for multi-mode research. In any case, we would prefer a more definitive test.
We knew we needed to keep the survey short in order to keep the mobile web respondents from dropping off but this was a very short 3.2 minute survey. Plausibly some of our findings may have been different had we used a survey length more common of web-based surveys (e.g. 15 minutes).

The above results lead us to the following conclusions:

- Based on this research, we should use horizontal rather than vertical formatting of rating scales.
- Short grids seem to shorten survey length and have no significant impact on abandon rates or data quality relative to grids broken into individual scale questions.
- The reduced acquiescence bias (the tendency to give agreeable answers) we have noted in mobile web relative to PC-web surveys appears to result from demographic and psychographic differences between respondents, not to the mode.
- Our test showed no measurable differences in the quality of data collected on PCs versus mobile devices.

Maritz Research has future research on research planned in this area and will be fielding a study with many of the same tests on a longer survey (we are targeting a 15 minutes survey). In addition to the scale and grid tests conducted here, we will also look into the impact design aspects of the survey such as background images and progress bars have on the respondent experience and data quality. We continue to refine our point of view around mobile web surveys.

Exhibit 1

<table>
<thead>
<tr>
<th>Cell</th>
<th>n</th>
<th>Mobile or PC</th>
<th>Scale Format</th>
<th>Grid or Separate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>202</td>
<td>PC</td>
<td>Horizontal</td>
<td>Grid</td>
</tr>
<tr>
<td>2</td>
<td>301</td>
<td>Mobile</td>
<td>Horizontal</td>
<td>Grid</td>
</tr>
<tr>
<td>3</td>
<td>201</td>
<td>PC</td>
<td>Horizontal</td>
<td>Separate</td>
</tr>
<tr>
<td>4</td>
<td>300</td>
<td>Mobile</td>
<td>Horizontal</td>
<td>Separate</td>
</tr>
<tr>
<td>5</td>
<td>201</td>
<td>PC</td>
<td>Vertical</td>
<td>Separate</td>
</tr>
<tr>
<td>6</td>
<td>300</td>
<td>Mobile</td>
<td>Vertical</td>
<td>Separate</td>
</tr>
</tbody>
</table>

The reduced acquiescence bias (the tendency to give agreeable answers) we have noted in mobile web relative to PC-web surveys appears to result from demographic and psychographic differences between respondents, not to the mode.
Appendix
Question Rendering on Mobile Devices
Horizontal and Vertical Scale Examples

Standard and Broken Grid Examples

For more information, please visit Maritz Research at www.maritzresearch.com or call (877) 4 MARITZ.