Do Smartphones Really Produce Lower Scores? Understanding Device Effects on Survey Ratings

Jamie Baker-Prewitt, PhD
Burke, Inc.

Abstract

As the proliferation of mobile computing devices continues, some marketing researchers have taken steps to understand the impact of respondents opting to take surveys on smartphones. Previously conducted research suggests a pattern of lower evaluative ratings from smartphone respondents, but whether the observed differences are truly driven by the data collection device or by characteristics of smartphone survey respondents themselves needs to be known. Leveraging the experimental control associated with a repeated measures research design, Burke conducted research to understand the implications of smartphone survey completion on the survey scores obtained. The findings suggest that the strength of previously found patterns is substantially diminished when respondent differences are fully controlled. This work reinforces the need to enable effective smartphone data collection for certain types of research among respondents whose smartphones are the data collection devices of choice.

Introduction

Predictably, smartphone ownership has continued to increase in the last year, with Pew reporting that 55% of US adults now own a smartphone (Pew Research Center, 2014). The marketing research industry has intentionally leveraged these mobile devices for in-the-moment data collection, sometimes asking respondents to take pictures and video or to complete a short survey during a shopping or product usage experience. Such mobile device use is often associated with qualitative research efforts, and it can provide rich consumer input that would be difficult to obtain without using respondents’ own mobile devices during data collection.

More recently, the industry has begun to understand the incidence and effects of unintentional smartphone data collection, where respondents use a smartphone to take a survey intended for a device with a much larger screen. Initial exploration of unintentional smartphone data collection uncovered some expected similarities and differences in data collected on different devices. Data quality has been shown to suffer when a smartphone is used for a survey and no attempt was made to adapt the survey for use on a small screen, but no device-based differences were observed in data on category purchase frequency or brand awareness (Baker-Prewitt and Miller, 2013). However, data from several pieces of research (Baker-Prewitt and Miller, 2013; Bartolone and Lattery, 2013; Peterson, 2013) as well as some client survey data examined to date evidence less frequent use of the first or most favorable response in brand ratings among respondents taking a survey on a smartphone. (See Figure 1 for one example.)
This tendency results in lower mean ratings as well. The pattern has been seen with different scale presentations, including...

- Different scale lengths (five, seven, ten or eleven points)
- Horizontal (grid) and vertical scale presentations

In addition, this phenomenon emerged in a study where respondents were told to use a smartphone to take the survey, but it has also been observed in situations where respondents themselves opted to use a smartphone to take a survey.

Furthermore, in various sets of data, applying weights to eliminate the demographic differences between computer and smartphone respondents attenuated the observed data differences, but this weighting did not completely eliminate the pattern of ratings differences observed. However, the fact that a device effect appeared to remain after weighting could be a function of not having the “right” demographic variables on which to weight.

Taken together, the previously conducted research suggests a pattern of lower evaluative ratings from smartphone respondents, yet the cause of this effect is not fully understood. Whether the observed differences are truly driven by the data collection device or by characteristics of smartphone survey respondents themselves needs to be known. Thus, Burke conducted additional research to understand the implications of respondent-driven smartphone survey completion on the data obtained.
Research Design

Burke conducted a repeated measures\(^1\) lab experiment to test for differences in data from surveys taken on a computer versus a smartphone. While a lab setting does not mimic typical settings in which respondents take surveys, it does provide a level of control necessary to infer a causal relationship between device usage and observed differences in survey ratings. Accordingly, through random assignment, half the respondents took a survey on a smartphone and then took a similar survey on a computer; the device order was reversed for the other half of respondents. All participants took the survey on both a computer and a smartphone.

Sampling

Research participants from the Cincinnati area were recruited to research sessions conducted at Burke headquarters. They met the following qualifications:

- Between 18 to 55 years old
- Own an iOS or Android smartphone and have a data plan
- Use the smartphone at least twice a week to access the Internet
- Do at least three activities (from a list) using the smartphone during a given week

The one-hour sessions were conducted in December and January. A total of 61 respondents participated in the research, and a rough balance was maintained between Apple and Android smartphone users and males and females.

Experiment

Multiple data collection sessions occurred in a lab setting. Upon arrival for a session, participants convened in a central room to receive a brief introduction, and then they were randomly assigned to go to one of two rooms. The room to which each participant was directed defined the order of data collection on each of two devices, specifically either smartphone and then computer, or computer and then smartphone. The experiment involved taking a survey on the first designated device in room one, participating in a filler task with all session participants in the central room from the introduction phase, and then taking a similar survey on the second designated device in a second room. (See Figure 2.)

\(^1\) The repeated measures design affords a high level of statistical power to detect significant differences in data produced from the two devices. The statistical power for measures examined in this research ranges from 70% to 90%; the range is driven by differences in base sizes and the magnitude of correlations between computer and smartphone data for different brand and customer measures.
Figure 2. Data Collection Sessions.

Questionnaire

About 70% of the survey that each participant took on both a computer and on a smartphone was identical, with 30% of the content varying to reduce the impact of taking a similar survey twice in a short time span. Smartphone surveys were adapted for ease of use and functionality on a small screen. Figure 3 shows the survey content and how the surveys differed.

Figure 3. Survey Content.
Research Findings

Demographic Profile

The infographic in Figure 4 shows the demographic profile of the research participants. In sum, slightly over half of participants are female, their average income is above the median US household income, their average age is almost 33, and nearly half has one or more college degrees.

Figure 4. Demographic Profile of Participants.

Category Purchase Frequency

As was measured in Burke’s 2013 work, this research examined how purchase frequency differed between responses obtained via computer and smartphone. The findings confirm those obtained in Burke’s previous work. Chi-square tests indicate that there are no device-driven differences in the frequency of purchasing sports drinks or visiting fast food or casual dining restaurants. (See Figure 5.)
In addition, respondents reported the amount of their average monthly cellular service bill, and no difference was found based on the data collection device (see Figure 6). This lack of effect is notable as respondents had quite different numerical data entry interfaces on a computer and a smartphone.
Brand Engagement

Statistical significance testing was performed to determine whether the data collection device affects measures of brand engagement when respondent differences are controlled through a repeated measures design. Specifically, device effects were tested for six brand engagement measures for each of six brands. (See the Appendix for images showing how these survey items were displayed.) Figure 7 shows that of the 36 tests performed, only one (3%) produced statistically significant results.

Figure 7. Summary of Brand Engagement Differences by Device.

The individual scales were measured on a ten-point, endpoint anchored scale.

Figure 8 shows the means obtained via computer and smartphone for two brand engagement measures for Gatorade. These results show relatively small differences in the means from the two devices.

Figure 8. Brand Engagement for Gatorade.
Figure 9 displays the magnitude of differences in the mean ratings obtained on a computer versus a smartphone for the six brands in three product categories. In these graphs, the numbers displayed represent the computer mean minus the smartphone mean. Negative numbers mean that the smartphone mean is higher than the computer mean; positive numbers indicate that the computer mean is higher than the smartphone mean.

Figure 9. Brand Engagement Differences between Computer and Smartphone Means.
For the one statistically significant difference found for AT&T as a trusted brand, the smartphone mean is higher than the computer mean. However, 72% of respondents either gave identical responses on the two devices, or their responses on the smartphone and computer differed by no more than one scale point on the ten-point scale.

With only one statistically significant difference found between the means from different devices, which is less than 3% of the 36 tests performed at the 90% confidence level, this set of findings suggests that no systematic device effect exists on the brand engagement measures.
Customer Engagement

Customer engagement measures were also measured for the same six brands in three product categories. The rendering of the customer engagement items differed from that of the brand engagement items (see the Appendix). As in Burke’s 2013 mobile research, these scale length and rendering differences are intentional to allow device effect testing for different types of survey items.

As done for the brand engagement measures, statistical significance testing was performed to determine whether the data collection device affects measures of customer engagement. Device effects were tested for five customer engagement measures for each of six brands. Figure 10 shows that of the 30 tests performed, three produced statistically significant results.

Figure 10. Summary of Customer Engagement Differences by Device.

![Customer Engagement Differences by Device](image)

The graphs in Figure 11 show the magnitude of differences in the mean ratings obtained on a computer versus a smartphone for the six brands in three product categories. As before, the numbers in these graphs represent the computer mean minus the smartphone mean, with negative numbers showing that the smartphone mean is higher than the computer mean, and positive numbers indicating that the computer mean is higher than the smartphone mean.

The Secure Customer Index™ can range from 0 to 100; the individual scales were measured on a five-point, fully anchored scale.
Figure 11. Customer Engagement Differences between Computer and Smartphone Means.

For Gatorade:
- Overall Satisfaction: SP > PC
- Likelihood to Purchase Again: SH > PC
- Likelihood to Recommend: SP > PC
- Earned Loyalty: PC > SP
- Preferred to Others: PC > SP

Five-point scales used. *Device difference significant at 90%.

For AT&T:
- Overall Satisfaction: -0.01
- Likelihood to Purchase Again: -0.01
- Likelihood to Recommend: 0.00
- Earned Loyalty: -0.01
- Preferred to Others: 0.00

Five-point scales used. *Device difference significant at 90%.
Interestingly, the three significant device effects found for customer engagement measures related to the brands in the fast food category. Figure 12 shows the frequency distribution for two of the measures on which a device effect appeared. For McDonald's, most of the effect is in the most favorable or top box rating, but for Taco Bell, both the top and second box rating evidence lower ratings from smartphones.
Figure 12. Frequency Distributions for Two Significant Devices Effects on Customer Engagement Measures.
Despite finding that on three of the 30 measures, smartphones yield significantly lower means compared to computers, the frequency distributions do not suggest that the differences are strong or managerially significant. This conclusion is further supported by examining the correlations between smartphone and computer responses for the three significant device effects on customer engagement measures. Figure 13 shows that the correlations are indisputably high.

Figure 13. Correlations between Computer and Smartphone Ratings.

Conclusions

The current research replicates the 2013 finding that responses to non-evaluative questions (in this case product category purchase frequency questions) do not evidence a data collection device effect. However, this research did not replicate the previously observed pattern of less favorable ratings from smartphone respondents.

- With a few exceptions, ratings from smartphones and PCs were not significantly different.
- In the four cases where the ratings did differ by device, three ratings were higher on the PC, and one was higher on the smartphone.

The significant differences that emerged are driven by a relatively small portion of respondents as 60% to 70% of respondents gave the same rating on a PC and smartphone. As further evidence that there is little effect of device on responses obtained, the correlations between respondents’ ratings on these items ranges from 0.81 to 0.84.
Limitations

No research is without limitations, and this work is no exception. As a lab experiment, the survey-taking experience of respondents in this experiment lacks mundane realism, which attenuates the external validity of the findings. While taking the surveys, research participants were in a room with other respondents taking the survey, and natural environmental distractions were not present.

Furthermore, despite inclusion of a qualitative filler task and various other distractions like physical movement between three different rooms, some respondents’ ratings might be the same on both devices because they remembered their responses from the first survey when they completed the second survey. Alternatively, some attempt to be consistent in responses might have influenced results.

In addition, the respondents were recruited and screened using agency lists of potential research participants in the Cincinnati area. Given that much survey research uses online panelists as respondents, replicating this research with panelists would be informative.

Implications

As ownership of smartphones and other devices for personal use continues to grow, so will the number of survey respondents who opt to use them to take surveys. This work suggests that researchers can support the appropriate use of smartphones for survey taking with little concern about device effects on survey results, i.e., measurement error.

In some cases, precluding smartphone data collection could have a biasing effect on the sample used for research by introducing non-coverage error. This potential bias would be particularly harmful when surveying populations that are more likely to elect to use a smartphone to take a survey or populations that are very small or hard to reach. The numbers below show the incidence of smartphone data collection for various populations targeted in Burke research:

- B2C ecommerce sample from a client: 15% used smartphone to take survey
- B2C Hispanic sample from a panel: 16% used a smartphone
- B2B specialty physicians from a panel: 18% used a smartphone
- B2C QSR consumers from a panel: 15% used a smartphone

In addition, not permitting smartphone data collection can also increase data collection costs and expand the time needed for data collection.

Observation of the research participants in this work suggests that simply advising respondents in the survey introduction to turn their smartphone to be landscape or horizontal instead of vertical might enhance the respondent experience. Many respondents did not turn their devices to a horizontal orientation, and interacting with the survey is visually easier when this simple step is taken.
Given the dearth of device effects found in this research, the patterns of differences that do exist between results from different devices are more likely driven by characteristics of the respondents who elect to use a smartphone for the survey or by some element of their product or service experiences. With one client data set, weighting smartphone respondents to mirror computer respondents demographically reduced but did not eliminate the device effect, but this finding could simply mean that the variables that drive different evaluative ratings were not available in the study.

While marketing researchers have spent years leveraging mobile devices for data collection, only recently have the effects of unintentional mobile data collection been uncovered and analyzed. With the incidence of smartphone survey completion established as substantial and expected to climb, researchers must understand whether the data collection device used impacts survey results. Given the removal of respondent differences in examining device effects in this research and the scarcity of device effects observed, this research provides assurance that researchers may permit smartphone data collection with little concern for biasing effects on the data obtained.
Appendix

Images of Brand Engagement Survey Items

Images of Customer Engagement Survey Items
References


Principal Author and Presenter

Jamie Baker-Prewitt, PhD
Senior Vice President, Director of Decision Sciences
Burke, Inc.
513.684.7663
jamie.baker-prewitt@burke.com

Jamie is Senior Vice President, Director of Decision Sciences at Burke, Inc. Jamie’s experience as a research consultant spans a wide variety of industries and a broad range of business issues. Jamie is a seminar leader for the Burke Institute and a speaker at industry conferences, including events sponsored by CASRO, AMA, SPSS, IIR, and ASQ. Jamie has published papers in Quirk’s Marketing Research Review and the CASRO Journal.

Jamie serves on Burke’s Board of Directors, Burke’s senior management committee, Burke’s R&D committee, and various other ad hoc management teams. Jamie led Burke through the process of obtaining ISO 20252 certification. She is a past president of the AMA Marketing Research Council, and she is immediate past Chair of the Advisory Board for the Masters of Science in Marketing Research program at the University of Texas at Arlington. Jamie is also a member of the American Psychological Association.

Jamie started her marketing research career at Burke in 1992. She holds PhD and MS degrees in social psychology from the University of Kentucky and a BA in psychology from Ohio University.