What Is The Impact of Smartphone Optimization on Long Surveys?

Shimon Sarraf
Jennifer Brooks
James Cole
Xiaolin Wang

Indiana University
Center for Postsecondary Research

Paper presented at the 2015 American Association for Public Opinion Research, Hollywood, Florida.

Introduction

Each year, an increasing number of college students are accessing online surveys using smartphones instead of personal computers. The National Survey of Student Engagement (NSSE), one of the largest annual North American college assessment projects with 1.5 million to 2 million undergraduate students, has seen exponential growth with its smartphone respondent population[1] (Sarraf, Brooks, & Cole, 2014). In 2011, only about 4% of its respondents used a smartphone, but by 2014 it had increased to 18%. Preliminary results from NSSE 2015 suggest this proportion has increased 50 percent from the prior year with roughly 27% using smartphones. The widespread adoption of smartphones among college students has naturally prompted data quality and survey format discussions and studies among some higher education researchers, especially as it relates to longer surveys like NSSE (Sarraf et al., 2014; Lambert & Miller, 2015). These discussions are prudent given degraded data quality often results when using an unoptimized survey format for a target population that relies on smartphones (Sarraf et al., 2014; Lambert & Miller, 2015). A number of helpful studies looking at smartphone survey optimization approaches have been completed recently based on relatively short surveys (Mavletova & Couper, 2014; de Bruijne & Wijnant, 2014) but more research is needed to inform smartphone optimization for longer surveys such as the most recent Buskirk and Andrus study (2014).

Using results from a ten-institution experiment using the 2015 National Survey of Student Engagement (NSSE) the current study details the impact that smartphone optimization has on a survey with over one-hundred questions. Study research questions center on how one long survey's optimization affected various data quality indicators, including completion rates, item nonresponse, duration, response option differentiation (straight-lining), and scale factor structures.

Background

Over the last two years, smartphone ownership has surpassed all other types of cell phones among adults in the US. In May 2011, only 35% of adult Americans owned a smartphone but by 2014, nearly two thirds (64%) possessed a smartphone (Pew Research Center, 2015). Duggan and Smith (2013) note that roughly one-third (34%) of smartphone users primarily access the internet with their phone. Though smartphone use is increasing, it is not the case that all American's have equal access to smartphones. A recent study indicates that smartphone ownership is stratified according to household income in the adult population. However, smartphone adoption is more evenly distributed among young adults (18-29 years old) (Smith, 2013). The Pew Research Center (2015) reported that by January 2014, 83% of 18-29 year olds owned smartphones. Hanley (2013) reported 92% of college students use smartphones to send and receive email messages, which may be particularly important for web-administered surveys that utilize email recruitment methods. Though the estimates are changing year-to-year, it is clear that use of mobile technology is pervading all aspects of young adults in the United States.

[1] The term "smarthphone" will be used throughout to indicate those using iPhones or any type of android phone device. This category does not include those using iPads, android tablets, or other larger screen devices.

A recent NSSE study (Sarraf et al., 2014) looked at college student characteristics by smartphone respondent status, finding many statistically significant but small differences between smartphone, tablet, and desktop computer respondents by gender, first-generation college status, standardized test scores, age, self-reported grades and race/ethnicity. Results showed smartphone respondents were more likely to have lower grades, be 30 years of age or older, and have lower standardized test scores. More meaningful differences, however, appeared within the smartphone respondent group. Compared to iPhone users, Android OS users were more likely to be first-generation college students, have lower standardized test scores, lower grades, and be a traditionally underrepresented minority.

With the widespread adoption of smartphone usage among college students, survey researchers now need to design the survey experience to accommodate mobile technology and respondent behaviors to facilitate maximum data quality. Earlier consensus on effective web survey design does not account for the increasing prevalence of mobile respondents. Ideally, new modes of administering surveys are tested rigorously before implementation, but the rapid consumer adoption of smartphones means that mobile respondents are steadily increasing even though there is no consensus on optimal design (Peytchev & Hill, 2010), especially as it relates to maintaining data quality. Assumptions are necessarily borrowed from previous studies on survey design but researchers seek empirical evidence that demonstrates how response quality may differ between mobile and non-mobile respondents. Peytchev and Hill (2010) engineered several tests to assess differences in data quality for mobile survey respondents. Randomizing response scales uncovered no bias between mobile and non-mobile respondents, nor did changing the order of questions. Other usability features common to mobile respondents, such as the smaller screen size, and differing navigational tools, such as physical keyboards or touchscreens, did adversely impact the quality of responses from mobile users (2010). For example, when it was necessary to scroll to see all response options, mobile respondents more often chose the first response value than did non-mobile survey participants. Findings from Stapleton (2013) illustrate similar results; mobile respondents more often select the response that can readily be seen even when the values of a satisfaction response scale are reversed. Stapleton also found mobile respondents abandon the survey more often than computer respondents, as did Maveltova in her 2013 study. Mavletova found no significant differences in primacy effect between mobile and computer respondents, however, nor were there differences between mobile and computer respondents when answering difficult or sensitive questions (2013).

An internal study analyzing data from the 2011 NSSE administration examined data quality from mobile respondents in several categories: survey drop off, item nonresponse, data mismatch between institution-reported and student-reported information, and a response quality indicator that aggregated three low-quality response criteria (Guidry, 2011). Guidry also found higher breakoff rates among smartphone users, though the other data quality indicators showed mixed results (2011). Using more current NSSE data, Sarraf et al. (2014) found that smartphone respondents differed in meaningful ways from personal computer and tablet users in terms of completion rates and missing data, though not in terms of various survey scale results. NSSE smartphone respondents were less likely to complete NSSE (62% versus 84%) and missing survey item percentages were far greater across the last two thirds of the survey, showing an approximate 20% point gap. A recent study by Lugtig and Toepoel (2015) focused on measurement error between PC, tablet and smartphone respondents. Using six panel surveys optimized for mobile devices, these researchers found that smartphone users had a greater proportion of missing items and lower survey experience evaluation scores. Based on the research to date, concerns regarding smartphone data quality appears warranted, thus making it prudent for survey designers to focus attention on optimal survey design.

Researchers have been experimenting with different optimization formats. One very recent study was conducted by DeBruijne and Wijnant (2014). These researchers tested several questionnaire design choices including scrolling versus paging, horizontal versus vertical answer scales, the number of vertically formatted response options, and open-ended versus closed ended questions. The results indicated that paging significantly increased duration time, however there was no significant difference in item nonresponse. Similarly, the study found no significant difference in missing items between horizontal and vertical answer scale layouts. In vertical layout, results also revealed that duration was significantly increased with more responses options. Follow-up analysis showed that 99% of respondents using vertical layout could see all five response categories, however that number dropped to 59% when viewing items with eleven response categories. This study also found evidence for primacy and recency effects where those presented with five response options were more likely to select the first response, and those presented with 11 response categories were more likely to select the last response. Mavletova & Couper (2014) also recently looked at differences between paging and scrolling formats, concluding “a scrolling design appears to make the process of completing the survey easier and more engaging for mobile web respondents (p. 509).”

Buskirk and Andrus (2012) detail three viable options to accommodate smartphone respondents. The do-nothing approach makes no special accommodation for mobile devices; the website simply displays as-is on the smaller screen, and the browser must scroll or navigate to view all content accordingly. Some college student surveys such as NSSE use this approach though the exact number is unknown. Another option requires development of a specialized app for the survey site. This approach is particularly effective at sizing images and survey content to a smaller-sized screen, but it may be cost-prohibitive because multiple applications (“apps”) must be developed for different operating systems. The app approach can also create a slower rate of advancement through the survey because each web page loads independently, which may frustrate users. A third option mimics the appearance of an app approach, but utilizes programming options (e.g., server side scripting and JavaScript) to enable a quicker load time for web pages. The web pages advance more quickly and appear more responsive than a non-mobile optimized version. This approach requires staff with sufficient programming skills, however, and can be compromised if a potential respondent has disabled JavaScript on their phone. Each of these approaches offer benefits, but none resolve all issues encountered by survey researchers. Among surveys aimed at college students it is currently unknown how many use the second and third approach. Buskirk and Andrus (2012) conclude there is no singular “right approach.” As Peytchev and Hill (2010) suggest, the best method of mobile optimization seems to be dependent upon the research project and sample composition. Survey length, question types, and response options may also influence a survey researcher’s perspective on the costs and benefits of the various approaches.

Study Rationale and Research Questions

It is still somewhat unclear from the literature how to best optimize a long survey like NSSE for smartphones. Common wisdom suggests surveys be shortened, however with established surveys this is not always a viable option. The current study set out to explore ways to improve the survey taking experience for smartphone users. Results are compared between NSSE smartphone optimized respondents, smartphone unoptimized respondents and those completing the survey on a desktop or laptop. More specifically, this study sought to shed light on the following questions:

1) Do student background characteristics vary among respondents based on the type of device they access the survey on? If so, what is the likely impact on other study results?
2) Does smartphone optimization for a long survey improve:
    a) early abandonment rates,
    b) completion rates,
    c) item nonresponse,
    d) survey duration, and
    e) straight-lining?
3) Does optimization for long surveys affect subjective evaluations of survey format?
4) Do assumptions of measurement invariance for ten NSSE scales hold?

Methods

Data Source & Sample. NSSE staff recruited ten US four-year colleges and universities to participate in a NSSE 2015 experimental administration aimed at better understanding the impact of smartphone optimization on survey data quality. All institutions had participated in NSSE 2014 and were known to have a high proportion of smartphone respondents. The experimental administration was identical to a standard NSSE administration, except for 50% of the sample having an opportunity to complete an optimized version of NSSE. Schools had five email recruitment messages sent on their behalf by NSSE to all first-year students and seniors. Institutions could decide to offer incentives, and, when they did, related text was included in all their recruitment messages. Each recruitment email included a URL that links to the informed consent statement and survey. The informed consent statement mentioned that the survey takes approximately 15 to 18 minutes to complete. Though we did not use available data for this study, institutions appended to the core NSSE instrument topical module item sets ranging in content from academic advising to civic engagement to information literacy.

Study institutions differed by Carnegie classification, total undergraduate enrollment size, and public-private status. The group included five research universities, four Master's colleges/universities, and one baccalaureate college. Two institutions had less than 5,000 total undergraduate students, four had between 5,000 and 10,000, and four had more than 10,000. Of the ten schools, two were privately controlled. Average institutional response rates (using AAPOR's RR2 calculation) for first-year students was 20% (with a range of 10% to 30%) and 24% for seniors (with a range of 13% to 35%).

Of the 38,245 first-year students and seniors included in the study sample, about half were randomly assigned to the smartphone optimized survey format group (n= 19,146) and the other half to the smartphone unoptimized survey format group (n=19,099). Students from both groups self-selected the device to complete the survey, whether that be a smartphone, personal computer or tablet. For the remainder of this study we refer to students that used the smartphone optimized survey version as the “optimized” group while those that used the smartphone unoptimized version as the “unoptimized” group. Respondents that either used a larger screen to complete NSSE, such as a desktop computer, laptop computer, or tablet, are grouped together under the generic term “desktop.”

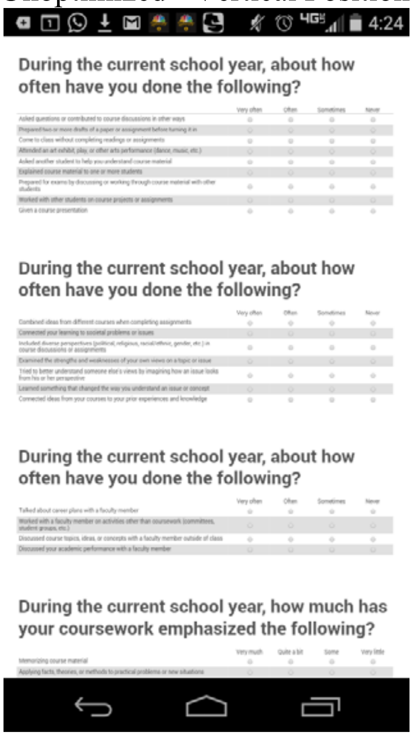
Of 7,735 respondents, we excluded 388 because of two issues. Due to an overlooked programming issue, 96 students completed an optimized survey format on their tablets despite the intention for all tablet users to receive the standard format. We also excluded another 292 respondents that switched in one way or another between a smartphone and larger screen device. This left a total of 7,347 respondents to analyze. Of this remaining population, 2,462 (34%) used their smartphones to participate while other students used either their personal computer or tablet. Of all smartphone respondents, 1,480 (60%) used the smartphone optimized survey version. Approximately 40% and 60% of all respondents were first-year students and seniors, respectively.

Survey Content and Format. The NSSE 2015 core survey included approximately 106 survey items (for a survey facsimile, see nsse.iub.edu/html/survey_instruments.cfm). Many of its survey items are organized into small groups that parallel the various content areas and scales NSSE uses in its research and reporting. Both the standard survey and the optimized version used for this study used a scrolling format, dividing items across four relatively long pages. NSSE survey items measure the extent to which college students engage in activities associated with various positive developmental outcomes: interacting with faculty, quantitative reasoning, collaborative learning, and other areas. The survey also asks students to reflect on how much their college emphasizes different things (spending significant time studying, providing academic support, etc.) and how they have developed in various areas (writing clearly, thinking critically, etc.). NSSE also asks respondents to provide background and demographic information. The majority of survey items use vague, four-point scales (e.g., very often, often, sometimes, never; very much, quite a bit, some, very little) while others use less vague response options, such as with items measuring hours spent preparing for class, participating in co-curricular activities, and working for pay.

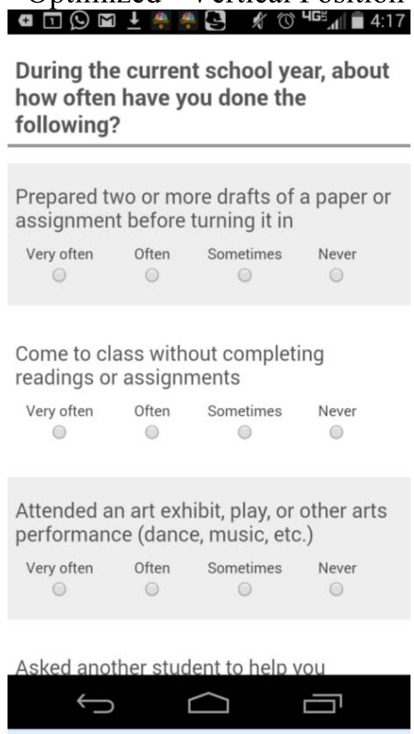
NSSE Smartphone Respondent Experience. Though tens of thousands of students have completed NSSE on their smartphones over the past several years, it is a less than optimal survey taking experience. Holding phones vertically, question stems are legible but specific questions are not, thus requiring respondents to zoom in on specific pieces of the question and response options to improve legibility (see Figure 1). By turning the phone horizontally, question stems and response options become much more legible though zooming may be required for those with less than perfect eyesight or whose eyes become fatigued looking at small text size. Due to finger size variation, for example, selecting response options accurately in either position may present challenges as well. The smartphone optimized version of NSSE attempted to address these various concerns by reformatting the many matrix tables. This was achieved by including horizontal response options underneath each individual matrix question, keeping the stem frozen at the top of the screen, and ensuring that zooming was not required to see text. To improve perceptions of progress, each completed item triggered a slight auto-advance so that the next unanswered question would be positioned in the middle of the screen.

Figure 1. Views of NSSE on Android Smartphone

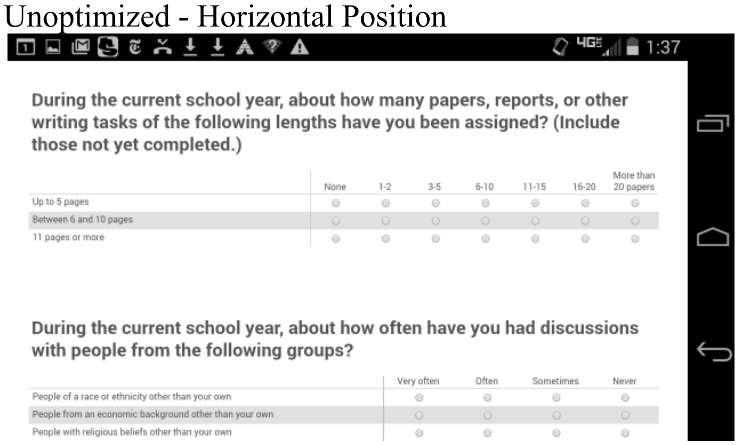
Unoptimized - Vertical Position



Optimized - Vertical Position



Unoptimized - Horizontal Position



Analyses. In order to assess the impact of smartphone optimization, we used several survey data quality indicators to look at how optimized smartphone respondents compared to both smartphone unoptimized and desktop groups. NSSE tablet respondent results traditionally mirror those of respondents using personal computers so these two groups were combined for this study (Sarraf et al., 2014). Given significant differences traditionally found between first-year students and seniors response rates, missing data, and survey responses, we analyzed both groups separately.

Though we randomly assigned sample members to either the optimized or unoptimized smartphone groups, not all students chose or had the ability to actually complete NSSE with a smartphone. In light of this unavoidable study design issue, and to strengthen our claims that smartphone optimization for a long survey does or does not matter, our first research question aimed to compare various demographic and background characteristics between the optimized and other groups to identify meaningful and statistically significant differences. Using Chi-square tests, the following variables were used to evaluate overall differences between groups: gender, age, race/ethnicity, parental education (Bachelor's degree or higher), cumulative grades, part-time enrollment status, and primary academic major. For race/ethnicity and major, we used a column proportions *z*-test with Bonferroni adjustment to determine significant proportional differences for particular background characteristic, such as Asian race or Social Sciences major. Given the unknown theoretical relationship between these characteristics and survey data quality indicators, we forgo multivariate analysis in favor of univariate statistical tests to answer our other research questions.

For the second research question about data quality, we first analyzed early abandonment rates to assess the impact of the optimized survey format. We defined students that abandoned the survey as those that clicked their email recruitment survey link, landed on the informed consent page, consented by clicking the "proceed to survey" button, and then chose not to answer any survey questions upon, presumably, assessing the first survey page. Chi-square tests determined statistically significant early abandonment rate differences between groups. We then analyzed missing data (a combination of both breakoff and item nonresponse behaviors) in three different ways: average percentage of missing items per student, percentage of respondents completing 100% of survey items, and percentage of respondents completing 95% or more survey items. To test for statistically significant differences, we ran four *t*-tests for the first measure and Chi-square tests for the second and third measures, each comparing optimized smartphone respondents to either unoptimized smartphone respondents or desktop respondents. As for all *t*-tests in this study, we tested for homogeneity of variance to determine the most appropriate p-value. We also included a line graph depicting missing data percentages for each survey item by the experiment's three different groups for both class levels to more clearly visualize the considerable variation that exists.

Related to missing data, we also reviewed the data for item nonresponse variation. We calculated the percentage of unanswered items for each respondent for each page. There were 33, 23, 31, and 17 items for the four pages. We then calculated the average percentage of item nonresponse by smartphone optimization status for each page. We used *t*-tests to indicate whether any statistically significant differences existed between groups.

To assess survey duration, we calculated the time in minutes it took for the approximately 45% of all survey respondents that completed all 106 survey items, using four *t*-tests to see if any observed time differences were statistically significant within either class level.

We analyzed the average number of items straight-lined on the first three survey pages. We define straight-lining (SL) as selecting the same response option for a set of items using the same scale. SL has been used as an indicator of *satisficing*, which is the process of "conserving time and energy and yet producing an answer that seems good enough for the purposes at hand" (Schaeffer and Presser, 2003, p. 68). There were potentially six, five and three sets of items to be straight-lined on the first three survey pages. Since no items shared the same response options on the fourth survey page, we excluded it from this analysis. *t*-tests results indicated whether any statistically significant differences existed.

At the end of the survey, we asked respondents two questions assessing the instrument's format:

1) Considering factors like ease of navigation, ease of reading the screen, and ease of selecting responses, please rate how easy it was for you to complete this survey. (response options: 1=Not at all easy to 7=Very easy)

2) How would you rate the visual design of the survey? (response options: 1=Poor; 2=Fair; 3=Good; 4=Excellent)

Besides using t-tests to look at the average scores between groups for both items, we also analyzed the proportion of respondents using the very top response options indicating either the survey was "very easy" to complete or that the visual design of the survey was "excellent" using Chi-square tests. Though one could reasonably argue that the survey completers responding to these questions at the end of the survey are not representative of the entire population since they do not include students that broke off, these results should be relatively reliable given that approximately 5,250 students (70%) provided feedback.

For projects like NSSE that rely on scales for various analytical purposes, knowing that smartphone optimization does not have any serious negative consequences is critically important. In order to shed light on this issue, we conducted multi-group confirmatory factor analysis (MG-CFA) for each NSSE scale, otherwise known as Engagement Indicators (EIs), in order to test measurement invariance across the three groups by class level. Confirming measurement invariance ensures that "psychometric test scores can be attributed to differences in the properties that such tests measure" and that a score relates "to the same set of observations in the same way in each group" (Borsboom, 2006). The ten EIs analyzed included: Higher-Order Learning (HO); Reflective & Integrative Learning (RI); Quantitative Reasoning (QR); Learning Strategies (LS); Collaborative Learning (CL); Discussions with Diverse Others (DD); Student-Faculty Interaction (SF); Effective Teaching Practices (ET); Quality of Interactions (QI); and Supportive Environment (SE). (For details about EIs, see nsse.iub.edu/html/engagement_indicators.cfm.) The MG-CFA for each EI followed several steps. First, a CFA was run separately for each group until the same model fit all groups well. If no model fit groups, measurement invariance was rejected and we pursued no additional testing. Second, assuming a model fit groups well, we then ran tests for configural, metric, and scalar invariance sequentially. Once a lower level of invariance was tested and rejected, we did not proceed with running tests for higher levels of invariance. Scalar invariance signifies the highest level of invariance, while configural is the lowest. Criteria used for determining acceptable model fit was RMSEA $<.06$, Chi-square p-value $>.05$, and CFI/TLI $>.90$. An even higher level of scalar invariance could be achieved when the Chi-square difference test p-values were greater than .05 and ΔCFI was less than .01.

Results

As shown in Table 1, though statistically significant differences across student background characteristics exist, in general, we did not see any sufficiently large differences between optimized respondents and the two other groups. We thus have confidence in drawing conclusions about the effect of optimization. Among the statistically significant differences found, the first-year optimized group had a proportionally greater number of students with parents holding a bachelor's degree or higher than the unoptimized group (55% versus 46%, $p<.05$) but a lower proportion compared to the desktop group (62%, $p<.01$). Compared to the desktop group, the first-year optimized group also had fewer students with self-reported grades of B+ or higher (56% versus 61%, $p<.1$) as well as differences by primary academic major ($p<.05$) and race/ethnicity ($p<.01$). Compared to the senior unoptimized group, senior optimized respondents included fewer older students (34% versus 41%, $p<.05$) and had a different distribution of primary academic majors ($p<.01$). Furthermore, compared to the senior desktop group, they had more females (68% versus 62%, $p<.01$), fewer students reporting at least one of their parents having a Bachelor's degree or higher (49% versus 56%, p.<.01), and fewer reporting B+ or higher cumulative grades (59% versus 68%, $p<.001$). We also saw differences in the distributions of race/ethnicity ($p<.01$) and academic major ($p<.001$).

Insert Table 1 about here

As Table 2 indicates, compared to the unoptimized groups, first-year and senior optimized smartphone respondents were much less likely to abandon the survey upon viewing the very first page of survey items. Only about 5% and 4% of first-year and senior optimized respondents, respectively, abandoned the survey early on before answering any questions, while 26% and 22% of the first-year and senior unoptimized groups did so ($p<.001$). Though the gap was not as great, meaningful differences with the desktop group existed as well ($p<.001$): approximately 12% of first-year students and 10% of seniors from the desktop group abandoned the survey upon viewing the first page of survey items.

Insert Table 2 about here

Smartphone optimization does also appear to reduce missing data though variation exists between first-year and senior populations. On average, first-year optimized respondents did not complete 27% of all NSSE items compared to 38% ($p<.001$) for the unoptimized group. Seniors showed a small, statistically insignificant 3% gap between these two groups (22% versus 25%). In contrast, desktop respondents had significantly fewer missing items than the optimized groups: 18% ($p<.001$) for first-year students and 14% ($p<.001$) for seniors. See Figure 2 for a look at missing data across the survey for the six groups.

Insert Figure 2 about here

As for survey completion rates, first-year students and seniors that used the optimized survey format showed meaningful and statistically significant differences compared to other groups. 49% of optimized first-year students completed all items but only 29% ($p<.001$) of the unoptimized group did so, a gap of 20%. Senior results were similar but not as pronounced (54% versus 41%; $p<.001$). Both optimized groups came out ahead of desktop respondents by around 7% ($p<.001$). These results varied somewhat notably when we defined completion by answering 95% or more of items. The 20% gap for first-year students decreased to 12% (61% versus 49%; $p<.001$) and we found no statistically significant difference for seniors (67% versus 63%). Comparisons between optimized and desktop groups showed the latter completing at higher rates: a gap of 11% for first-year student and 12% for seniors ($p<.001$) existed.

Results for item nonresponse, defined for this study as respondents intentionally or unintentionally skipping items, showed both meaningful and statistically significant variation between groups though not in any consistent manner across the four pages. For the first page, approximately 9% of the 33 items went unanswered by first-year optimized respondents versus 19% ($p<.001$) and 9% (n.s.) for unoptimized and desktop groups, respectively. A somewhat similar pattern, albeit less notable, emerged for seniors where optimized and unoptimized respondents did not complete 6% and 11% of first-page items; the optimized group showed virtually identical results to the desktop group. Second page results indicated almost the same proportion of item nonresponse between the two smartphone groups for both class levels, though the two desktop groups had about half the amount compared to their respective optimized groups ($p<.001$). For the third page, first-year optimized respondents had on average 5% item nonresponse versus 12% ($p<.01$) for those that used the unoptimized survey format; we saw no notable difference for seniors however. No first-year difference with the desktop group was found though senior optimized respondents had a statistically significant result (6% versus 3.5%; $p<.05$). The fourth page showed practically no item nonresponse across the groups with a 1% average.

Optimization appears to have substantially reduced survey completion time for both classes. First-year and senior optimized respondents completed all survey items in 12.2 minutes compared to 15.0 ($p<.001$) and 14.6 ($p<.001$) minutes for unoptimized groups. These differences represent an approximate 16% to 19% reduction in completion time for smartphone respondents. Somewhat surprisingly, we see optimized groups taking about .8 minutes less time than their respective desktop groups ($p<.1$ for first-year students and $p<.05$ for seniors).

Overall, optimized respondents straight-lined less frequently than unoptimized respondents across both class levels, but less so with desktop respondents. For the first page with 6 sets of items that could be straight-lined, first-year and senior optimized groups straight-lined 1.4 and 1.6 item sets compared to 1.8 ($p<.01$) and 1.9 ($p<.05$) for unoptimized groups, respectively. We found similar statistically significant results when making comparisons to desktop respondents. For the second page with 5 sets of items, we saw fewer statistically significant results though optimized groups consistently showed less straight-lining in relation to unoptimized groups (first-year students: 1.0 versus 1.3, $p<.01$; seniors: 1.1 versus 1.3, $p<.05$). Optimized groups showed no meaningful differences with desktop respondents. On the third page with only three potential item sets, we found between .3 to .5 sets straight-lined with the only statistically significant difference exhibited between senior optimized and unoptimized groups (.3 versus .4, $p<.05$).

Questions assessing completion ease and survey design indicated that smartphone optimization improves the respondent experience. Regarding ease of completion, both optimized groups had an average score of 6.2 out of 7 ("very easy") compared to an average of 5.8 ($p<.01$) and 6.0 ($p<.05$) for the first-year and senior unoptimized groups. Notably, we found no statistically significant differences with the desktop group. More specifically, approximately 60% of first-year and senior optimized respondents marked the highest response option. Only 46% ($p<.01$) of the first-year unoptimized group and 52% ($p<.05$) of the senior unoptimized group did so. In terms of the visual design, students generally felt NSSE's design was good with first-year and senior average result patterns looking almost identical, ranging between 3.1 and 3.3 out of 4 ("excellent"). The optimized group came out ahead of the two other groups by slim margins; three out of the four statistical tests showed differences to be statistically significant (the senior comparison with the desktop group was the one exception). About a third of all respondents reported the survey's design to be "excellent" with somewhat limited variation across groups. First-year optimized respondents showed both meaningful and statistically significant differences with the two other groups: a 14% gap ($p<.001$) with unoptimized respondents and a 7% gap ($p<.01$) with desktop respondents. No statistically significant differences could be found between senior groups.

In terms of EI measurement invariance across the three groups, all first-year and senior EIs met the criteria for scalar invariance except for Learning Strategies (see Table 3). Only Quantitative Reasoning and Collaborative Learning for first-year students met the highest measurement invariance criteria though, while for seniors Discussions with Diverse Others and Effective Teaching Practices did. For more detailed results, please contact the authors.

Insert Table 3 about here

Discussion

Results validate that smartphone optimization improves survey data quality even while ignoring prevailing wisdom that surveys be shortened to accommodate smartphone respondents. More specifically, we demonstrated that long survey optimization efforts are worthwhile in terms of early abandonment, missing data, item nonresponse, duration, straight-lining and subjective evaluations of the survey's appeal. This study also made clear that optimization does not address NSSE's missing data issue entirely, despite the noticeable improvements. We also learned a few things about smartphone survey design. For instance, somewhat unconventional design choices, such as maintaining radio buttons instead of creating larger thumb-friendly buttons, do not negate optimization efforts.

Interestingly, results for first-year students are more pronounced than for seniors when it comes to missing data. We do not know why optimization appears to be more impactful for younger students in this regard. Given increasing rates of market penetration of smartphone adoption, particularly among younger demographics, is it possible that the younger students have greater intolerance for unoptimized web content? Were they earlier adopters of smartphone technology and thus have greater expectations? We don't know the answers to these questions, but the disparity between younger and older students on this measure is worth tracking in future studies.

Our average item nonresponse results provide additional insights about survey pages that benefit more or less from optimization. Though results were a bit mixed by page and class level, overall, optimization does appear to reduce the incidence of either intentional or accidental item skipping. In fact, the optimized group actually ended up looking more similar to desktop respondents than unoptimzied respondents. While results from this study are satisfactory, item nonresponse among smartphone respondents still exceeds desktop respondents. It may be worthwhile to slightly revise design choices for future studies in order to bring item nonresponse percentages into even greater equilibrium with desktop respondents. Additionally, because of NSSE's length, the longer completion duration of smartphone respondents has been a concern, but the optimized format we used resulted in significantly lower duration, even when compared to desktop users.

Assessment questions at the end of survey confirm that the optimized version is better received by respondents. Comments from an open-ended comment box showed a preference for more color on the survey, but smartphone optimized respondents still reported the survey was "very easy" to complete and was on par with the desktop group. Of course, one caveat with this finding is that the opinions of students that brokeoff from the survey were not included. Our feeling is that the results for the optimized and other groups are most likely reliable, but unoptimized results may overestimate the somewhat surprisingly positive results for this group.

We expected optimization to improve the experience of smartphone respondents. What we did not expect, however, was that smartphone optimized respondents would equal or surpass the desktop group on several measures. For instance, achieving parity in completion time was incredibly important for a long survey where many participants complain about its length. In addition to shorter duration times, proportionally more smartphone optimized respondents completed every survey item and fewer abandoned the survey upon viewing the first survey page. Similarly, first-year smartphone optimized survey respondents indicated the enhanced design may provide a better user experience than the desktop format.

Peytchev and Hill (2010) have emphasized that in the absence of a uniform set of best practices optimization efforts should be designed with the goals of the research project and sample in mind. Several considerations guided design choices for the optimized NSSE instrument. First, there was little interest in shortening survey content to accommodate smartphone respondents. Student engagement is not a single construct, but rather a collection of ideas about activities that are educationally purposeful that together constitute "engagement." Shortening survey content would certainly exclude important activities and compromise the utility of NSSE results for participating colleges and universities. In addition, because EIs feature prominently in NSSE results optimization needed to preserve scale properties. To date, optimization's impact on scale properties has not received much attention in the literature even though measurement invariance is extremely important for assessing differences between groups. Many institutions track their EI scores overtime so we must ensure that score changes are not influenced by optimization choices. That the Learning Strategies EI does not meet criteria for measurement invariance is a concern and an area that NSSE staff will surely need to discuss more before any full-scale survey optimization is implemented. Lastly, given NSSE's many question matrices, the biggest optimization challenge may have been reformatting all the matrix questions and response options into something easily read on a smartphone while minimizing perceptions of length. Freezing question stems at the top of smartphone screens as scrolling occurs, along with a slight auto advance upon answering questions, seems to have effectively addressed these concerns, thus providing a sensible design template for long (or short) surveys.

In a climate of generally declining survey response, survey fatigue is a hot topic among college and university administrators that depend on assessment data, and there is a perception (whether real or imagined) that students have become anesthetized to invitations to participate in surveys or are unwilling to complete them once begun. When NSSE began surveying students in 2000, surveys were not as prevalent on college campuses compared with now due in large part to the prevalence of "out of the box" survey sites like SurveyMonkey and Qualtrics. Though helpful to many, their arrival has undoubtedly contributed to survey fatigue and degraded data quality for long surveys such as NSSE. Of course, we should also recognize there is a good reason for all the survey activity. Colleges and universities face increased pressures to provide evidence that they are delivering on their promises, whether related to curricular or co-curricular offerings. Given this environmental context and our findings, college survey administrators' smarthphone optimization efforts should pay dividends, especially for response rates at institutions particularly saturated with smartphone users. Our early abandonment findings suggest that an additional 20% of smartphone consenters will turn into formal survey respondents once NSSE provides all smartphone users with an optimized survey format.

Overall, optimization efforts with the NSSE 2015 administration demonstrate that a long survey can be successfully optimized while still maintaining the validity of most scales. Future efforts will likely build upon this success by enlivening the survey format with color and testing other design variations.

References

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, *44*, 176-181.

Buskirk, T. D., & Andrus, C. (2012). Smart surveys for smart phones: Exploring various approaches for conducing online mobile surveys via smartphones. *Survey Practice*, 5. Available at: <http://36.22.354a.static.theplanet.com/index.php/SurveyPractice/article/view/63>. Date accessed: May 12, 2014.

Buskirk, T. D., & Andrus, C. (2014). Making mobile browser surveys smarter: Results from a randomized experiment comparing online surveys completed via computer or smartphone. *Field Methods,26* , 322-342.

DeBruijne, M. & Wijnant, A. (2014). Improving response rates and questionnaire design for mobile web surveys. *Public Opinion Quarterly, 78*, 951-962.

Duggan, M., & Smith, A. (2013). *Cell Internet Use 2013*. Pew Research Center, Washington D.C. Available at: http://pewinternet.org/Reports/2013/Cell-Internet.aspx. Date accessed: March 28, 2014.

Guidry, K. R. (2012). *Response Quality and Demographic Characteristics of Respondents Using a Mobile Device on a Web-Based Survey*. Paper presented at AAPOR Annual Conference, Orlando, FL.

Hanley, M. (2013). *College Student Smartphone Usage Hits 74%; Tablet Ownership at 30%.* Available at: http://th-ebooks.s3.amazonaws.com/Hanley_Ball_State_College_Student_Cell_Phone_Study_Summary_February%202013.pdf. Date accessed May 1, 2014.

Lambert, A. D. & Miller, A. L. (2015). Living with smartphones: Does completion device affect survey responses. *Research in Higher Education*, *56*, 166-177.

Lugtig, P, & Toepoel, V. (2015). The use of PC's, smartphones, and tablets in a probability-based panel surveys: Effects on survey measurement error. *Social Science Computer Review*. DOI: 10.1177/0894439315574248.

Mavletova, A. (2013). Data quality in PC and mobile Web surveys. *Social Science Computer Review, 31,* 725-743.

Mavletova, A. & Couper (2014). Mobile web survey design: Scrolling versus paging, SMS versus e-mail invitations. *Journal of Survey Statistics and Methodology*, 2, 498-518.

Pew Research Center (2015). *U.S. Smarpthone Use in 2015*. Retrieved from http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/National Survey of Student Engagement (2013). *NSSE 2013 Engagement Indicators*. Center for Postsecondary Research, Indiana University. Available at: http://nsse.iub.edu/2013_institutional_report/pdf/EI%20Intercorrelations%202013.pdf. Date accessed: April 25. 2014.

Peytchev, A, & Hill, C. (2010). Experiments in mobile web survey Design: Similarities to other Modes and Unique Considerations. *Social Science Computer Review, 28*, 319-335.

Sarraf, S., Brooks, J., & Cole, J. (2014). *Taking surveys with smartphones: A look at usage among college students.* Paper presented at AAPOR Annual Conference, Anaheim, CA.

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, 29, 65-88.

Smith, A. (2013). *Smartphone Ownership 2013*. Pew Research Center, Washington D.C. Available at: http://www.pewinternet.org/Reports/2013/Smartphone-Ownership-2013.aspx. Date accessed on March 28, 2014.

Stapleton, C. (2013). The Smart(Phone) way to college survey data. *Survey Practice*. 6 (2). Available at: http://surveypractice.org/index.php/SurveyPractice/article/view/75/html. Date accessed: March 28, 2014.

The American Association for Public Opinion Research (2015). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 8th edition*. AAPOR: Washington, DC.

Table 1. Student characteristics by smartphone optimization status and class level

	All	First-Year Students			Seniors		
		Optimized	Unoptimized	Desktop	Optimized	Unoptimized	Desktop
Female	65	67	69	67	68	66	62**
Age[a]							
24 years old or greater	24	4	3	4	34	41*	36
Race/Ethnicity[a]							
American Indian or Alaska Native	1	1	.0	1	1	1	.4
Asian	3	3	3	3	3	2	3
Black or African American	11	15	16	9*	14	14	9*
Hispanic or Latino	6	6	9	6	6	6	5
Native Hawaiian or Other Pacific Islander	.0	.0	.4	.4	.0	.0	.4
White	63	58	51	62	59	62	66*
Other	2	2	1	1	2	1	2
Multiracial	7	6	8	9	6	5	7
I prefer not to respond	4	3	5	4	6	4	4
International	4	6	7	4*	3	4	3
Parental Education[a]							
Bachelor's or higher	56	55	46*	62**	49	48	56**
Cumulative Grades[a]							
B+ or higher	63	56	55	61+	59	63	68***
Part-time enrollment	13	4	5	4	21	19	18+
Primary Academic Major[a]							
Arts & Humanities	10	5	10*	9*	12	8*	11
Biological Sciences, Agriculture, & Natural Resources	13	15	14	15	14	10	13
Physical Sciences, Mathematics, & Computer Science	5	6	8	7	3	5	5*
Social Sciences	14	10	9	14	12	12	16*
Business	15	12	13	13	15	18	16
Communications, Media, & Public Relations	4	3	4	5	4	5	5
Education	7	9	7	7	8	10	7
Engineering	7	10	7	8	6	7	8
Health Professions	13	19	20	14*	17	17	10*
Social Service Professions	4	5	4	4	5	4	3*
All Other Majors	5	4	1	4	6	4	7
Undecided, undeclared	1	3	4	2	.0	2*	.1

[a] Self-reported survey information. Results based on approximately 5,400 respondents. Gender and part-time status statistics used all survey respondents (n=7,347).
+ p< .1; * p< .05; ** p<.01; *** p<.001

Table 2. Survey data quality outcomes by smartphone optimization status and class level

	All	First-Year Students			Seniors		
	(n=7,347)	**Optimized** (n=727)	**Unoptimized** (n=433)	**Desktop** (n=1,798)	**Optimized** (n=753)	**Unoptimized** (n=549)	**Desktop** (n=3,087)
Early abandonment rate (no statistical test results available)	11.5%	4.6%	25.8%***	11.9%***	4.2%	21.9%***	9.6%***
Average percentage of missing items per student	19%	27%	38%***	18%***	22%	25%	14%***
Percentage of respondents completing 100% of survey items	45%	49%	29%***	43%**	54%	41%***	46%***
Percentage of respondents completing 95% of survey items	71%	61%	49%***	72%***	67%	63%	79%***
Average item nonresponse							
Page 1 (33 items)	8.5%	9.1%	19.2%***	8.6%	6.2%	10.8%**	6.9%
Page 2 (23 items)	5.0%	8.7%	9.8%	4.2%***	7.2%	7.9%	3.3%***
Page 3 (31 items)	4.7%	4.9%	11.8%**	4.1%	6.2%	7.7%	3.5%*
Page 4 (17 items)	1.3%	1.1%	2.6%	1.5%	1.6%	1.3%	1.2%
Average completion time (in minutes; for those without missing items)	13.0	12.2	15.0***	13.0+	12.2	14.6***	12.9*
Straight lining scale items							
Page 1 scales (out of 6 scales)	1.8	1.4	1.8**	1.7***	1.6	1.9*	1.9***
Page 2 scales (out of 5 scales)	1.1	1.0	1.3**	1.1	1.1	1.3*	1.1
Page 3 scales (out of 3 scales)	.3	.3	.5	.3	.3	.4*	.3
Assessment Question #1 [a]	6.2	6.2	5.8**	6.1	6.2	6.0*	6.3
% reporting "Very easy"	57%	59%	46%**	55%	61%	52%*	58%
Assessment Question #2 [b]	3.2	3.3	3.1**	3.2**	3.3	3.1*	3.2
% reporting "Excellent"	36%	41%	27%***	34%**	40%	35%	37%

+ p< .1; * p< .05; ** p<.01; *** p<.001

[a] Assessment question #1: Considering factors like ease of navigation, ease of reading the screen, and ease of selecting responses, please rate how easy it was for you to complete this survey. Response options: 1=Not at all easy to 7=Very easy

[b] Assessment question #2: How would you rate the visual design of the survey? Response options: 1=Poor; 2=Fair; 3=Good; 4=Excellent

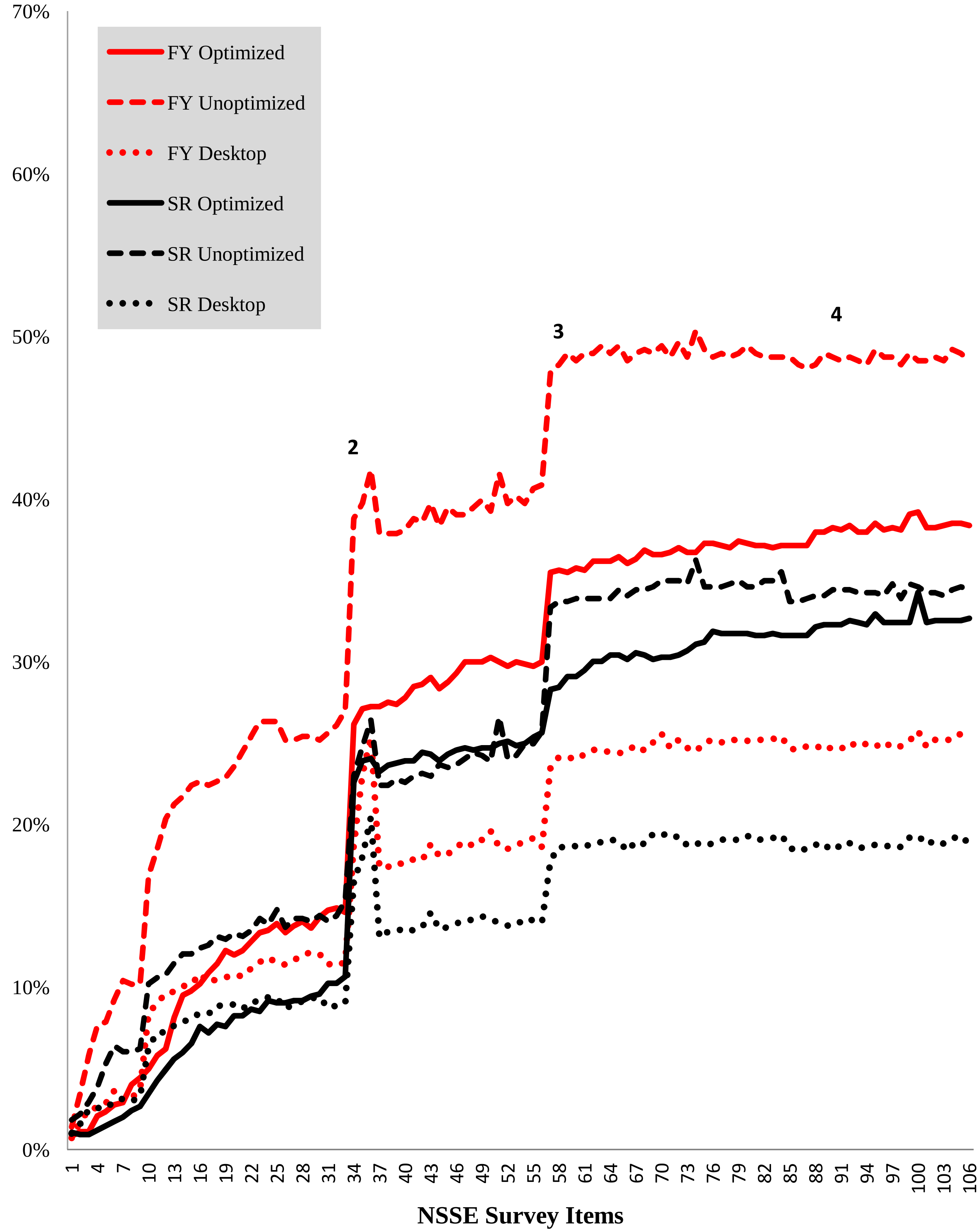
Note: Approximately 5,250 completed the assessment questions at the end of the survey.

Table 3. Multi-group invariance results by Engagement Indicator

	Students	Seniors
Higher-Order Learning	scalar	scalar
Reflective and Integrative Learning	scalar	scalar
Quantitative Reasoning	scalar +	scalar
Learning Strategies	*variant*	*variant*
Collaborative Learning	scalar +	scalar
Discussions with Diverse Others	scalar	scalar +
Student-Faculty Interaction	scalar	scalar
Effective Teaching Practices	scalar	scalar +
Quality of Interactions	scalar	scalar
Supportive Environment	scalar	scalar

Note: "scalar" implies the scale achieved strong measurement invariance as indicated by *most* model fit indices. "scalar +" implies that strong measurement invariance is supported by *all* model fit indices. "*variant*" implies the scale did not achieve measurement invariance since configural invariance was supported by less than half of model fit indices.

Figure 2. Missing item rates by smartphone optimization status and class level



Note: Numbers 2, 3, and 4 represent the beginning of the second, third and fourth survey pages.