Grid Formats, Data Quality, and Mobile Device Use:

Toward A Questionnaire Design Approach
Understanding and approaching mobile survey taking

“The control which designers know in the print medium, and often desire in the web medium, is simply a function of the limitation of the printed page. We should embrace the fact that the web doesn’t have the same constraints, and design for this flexibility. But first, we must ’accept the ebb and flow of things.’”

John Allsopp, “A Dao of Web Design”
Mobile + tablet use over time

- Across several multi-institutional studies of college students, we have observed smartphone + tablet usage to access the surveys roughly double each year.
- For similarly designed studies we’ve fielded in 2013, usage is hovering around 10%.

Smartphone + Tablet Access Rates, 2009-2013
Considerations for addressing the mobile space

- Population/rates of mobile device use
- Questionnaire integrity (previous administrations, scales)
- Researcher tolerance for change
- Financial and time constraints
- Emerging body of research on mobile survey taking
  - Approaches by Callegaro (2010) and Buskirk and Andrus (2012)
  - Much descriptive research; recent experiments
- Technical factors
  - Respondent-side: Device variation, environment, etc.
  - Programmer-side: Technical constraints of survey system

- These last two items represent this “ebb and flow of things” and are evolving quickly, necessitating a return to the beginning: solid, and possibly adapted, questionnaire design.
Grids and mobile browsers

- Frequently examined by researchers, used to display items in a space efficient way... on a computer

- Potentially even more problematic on a mobile device because:
  - Screen space is at a premium
  - Thumb positioning may make it difficult to differentiate
  - Sensitivity of touch screens may lead to respondent errors

- We’d hypothesize that responses on smartphones are particularly susceptible to variable factors:
  - Differences in mobile devices and browsers (compatibility)
  - Capabilities of survey systems to handle these formats
  - Questionnaire design factors (context, burden, etc.)
Breakoffs: Grids and questionnaire context

Post-Consent Breakoffs in a Multi-Institutional Study of College Study Mental Health

Observed same patterns on other multi-institutional studies of college students
Non-differentiation: Grids and questionnaire context

- Six grids with same Likert scale (39 items total); occasional items reverse coded
- Core measures of 2012 Multi-Institutional Study of Leadership
- High levels of straightlining; even higher for those completing the survey on mobile device (McClain et al, 2012)
  - Mobile users also more likely to straightline to the far left
  - Possible indication of satisficing (Krosnick, 1991)
Case study: Grids, mobile devices, and programmatic decisions
Exploring grid interaction on mobile devices

- Study of undergraduate students at a large Midwestern university (biannual since 1999)
  - Drug/alcohol use, sexual activity, and other student behavior
  - Web survey; email prenotification, invitation, and reminders
  - Overall: 34.41% AAPOR RR2, 16.05% breakoff rate, 10.34% smartphone access rate

- 2013: Implementation on two systems (Rs randomly assigned to primary system or secondary “pilot”; n = 10,000, 2,000)
  - Can we essentially equalize the R experience across systems...
  - And if so, does some sort of “system effect” still exist?
  - Same data collection protocol, slightly different contact schedule
  - Exploratory analysis of paradata, substantive responses underway
• While we had thus far been using what Buskirk and Andrus (2012) call a Passive-Mobile Browser Approach on our primary system...

• In programming the second survey, we tested the pilot system’s Active-Mobile Browser Approach—including grid “optimization.”
  ➢ Driven by necessity: wide grids, survey design elements
  ➢ But, also a prime opportunity to further explore one way we might alleviate burden that grids place on mobile respondents.

• Not a perfect experimental design; but a prime example of collisions of implementation challenges and questionnaire design issues that arise in the mobile space.
The alcohol self-harms grid: iPhone (Portrait Orientation)

Primary (Passive) Implementation

Pilot (Active) Implementation
The substantive measures

• “Please indicate how often **during the past 12 months** you have experienced the following as a result of drinking.”
  - Only asked of those who drank in past 12 months
  - Five pt. scale from “no occasions” to “10+ occasions,” + RNS
  - Range in severity/level of harm

• Many of the qualities that our exploratory analyses have suggested to be of importance:
  - Early in questionnaire
  - Sensitive content
  - Fairly complex recall task
  - Two grids in close proximity to each other, same scale
  - Decent variety of harm levels/expected reporting
Is the grid problematic to begin with?

- Examined key indirect indicators of data quality within each system, for computer Rs (isolated from design tweaks)

- **Breakoffs:** Peaked around this grid; of those who progressed past the consent question...
  - 30.9% (primary) / 26.4% (pilot) of remaining breakoffs occurred either on or immediately after this grid

- **Straightlining:** 28.4% overall, and almost all to the far left
  - 28.1% (primary) / 30.0% (pilot)

- **Item missing data:** Very little overall

- **Substantive responses:** Distribution looks as expected

- Within the standard/primary grid implementation, surprisingly not much difference between mobile and computer Rs
Is there an independent effect of the system?

Among computer Rs only...

- **Breakoffs:** No significant difference in overall rates (11.5% primary, 9.9% pilot)
  - Pilot: more consent breakoffs (41.4%/21.9%, $X^2=3.299$, $p=.091$)
  - After consent question, similar breakoff patterns

- **Elapsed time:** Significant, practically small diff. in means
  - Primary system 19.00, pilot system 18.02

- **Straightlining:** Few sig. differences; no overall system trend. Didn’t seem to correlate with substantive response patterns

- **Substantive responses:** 2 of 25 items on two key grids favor the primary system with higher reporting
  - If collapsing down to “any" reporting, rises to 6 of 25

- Many avenues for further analysis, but overall a somewhat comparable experience
The alcohol self-harms grid: Computer

<table>
<thead>
<tr>
<th>Community</th>
<th>About You</th>
<th>Tobacco &amp; Alcohol</th>
<th>Drugs</th>
<th>Sex &amp; Relationships</th>
<th>Future Plans</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pilot System</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Standard System</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please indicate how often during the past 12 months you have experienced the following as a result of drinking.

<table>
<thead>
<tr>
<th>Behavior</th>
<th>No occasions</th>
<th>1-2 occasions</th>
<th>3-6 occasions</th>
<th>6-9 occasions</th>
<th>10+ occasions</th>
<th>Rather not say</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performed poorly on a test or important project.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missed a class or work due to drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driven a car while under the influence of alcohol.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driven a car after drinking 5 or more drinks in a two-hour period.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Been hurt or injured after drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vomited.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Been taken advantage of sexually.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Took advantage of another sexually.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seriously thought about suicide.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were afraid you might be alcoholic.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Been annoyed by people criticizing your drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Had a drink first thing in the morning as an “eye opener” or to get rid of a hangover.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Felt guilt or remorse after drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Felt that you should cut down your drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Had unplanned sex.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Had blackouts (amnesia).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please indicate how often during the past 12 months you have experienced the following as a result of drinking.

<table>
<thead>
<tr>
<th>Behavior</th>
<th>No occasions</th>
<th>1-2 occasions</th>
<th>3-6 occasions</th>
<th>6-9 occasions</th>
<th>10+ occasions</th>
<th>Rather not say</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performed poorly on a test or important project.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missed a class or work due to drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driven a car while under the influence of alcohol.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driven a car after drinking 5 or more drinks in a two-hour period.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Been hurt or injured after drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vomited.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Been taken advantage of sexually.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Took advantage of another sexually.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seriously thought about suicide.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were afraid you might be alcoholic.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Been annoyed by people criticizing your drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Had a drink first thing in the morning as an “eye opener” or to get rid of a hangover.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Felt guilt or remorse after drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Felt that you should cut down your drinking.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Had unplanned sex.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Had blackouts (amnesia).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Could an Active Browser approach for smartphones improve response quality?

- Pilot smartphone R’s:
  
  - Completed with a **median lower elapsed time** (20 min vs. 23 min); no difference in means
  
  - **Broke off less frequently overall** than primary smartphone Rs (32.3% vs. 37.1%), though not significant
  
  - **Broke off more frequently at consent question**; similar patterns to primary afterwards
  
  - Left **very few items blank** (similar across all treatments)
  
  - **Straightlined less frequently** on the self-harms grid (10.5% of R’s vs. 25.5%, p=.059).
Substantive response differences among smartphone Rs

- For 11 of 25 items, using the **pilot system** resulted in **higher reporting** of these somewhat sensitive behaviors in initial regressions.

- Could reflect higher motivation (earlier breakoffs in pilot system), but we have initially found few differences in surrounding questions about drinking behaviors (select one, select all, numeric formats).

- Along with system, controlled for gender, ethnicity (white/nonwhite), class, age (traditional/not).

<table>
<thead>
<tr>
<th>Item</th>
<th>B (SE): Pilot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor performance on test/project</td>
<td>1.545 (.416)***</td>
</tr>
<tr>
<td>Missed class or work</td>
<td>1.192 (.377)***</td>
</tr>
<tr>
<td>Seriously thought about suicide</td>
<td>1.087 (.645)*</td>
</tr>
<tr>
<td>Been annoyed by criticism</td>
<td>0.773 (.424)*</td>
</tr>
<tr>
<td>Had drink first thing in morning</td>
<td>1.008 (.546)*</td>
</tr>
<tr>
<td>Felt guilt/remorse after drinking</td>
<td>1.143 (.371)***</td>
</tr>
<tr>
<td>Felt should cut down drinking</td>
<td>0.666 (.375)*</td>
</tr>
<tr>
<td>Had unplanned sex</td>
<td>0.675 (.402)*</td>
</tr>
<tr>
<td>Had blackouts (amnesia)</td>
<td>1.078 (.367)***</td>
</tr>
<tr>
<td><strong>Harms from other:</strong> Sexually assaulted by someone drunk</td>
<td>1.219 (.591)**</td>
</tr>
<tr>
<td><strong>Harms from other:</strong> Rode with drunk driver</td>
<td>1.190 (.425)***</td>
</tr>
</tbody>
</table>

*** p<.01, ** p <.05, * p<.1
Substantive response differences among smartphone Rs

Percentage of smartphone respondents reporting self-harms

Full distribution of “Felt guilt or remorse after drinking” (B14a)
Why a design focus?
What we learned

- Not a perfect design (small n’s, system variation); but a very practical example of the complex decisions many of us face

- Drawing from exploratory analysis, some effects of the survey system, even with no format changes for computer Rs
  - Some impact on breakoffs and, in one case, substantive responses
  - Extremely relevant for studies that switch survey systems, in terms of research generalizability
  - More detailed analysis in progress

- In this case, grids that broke the items apart increased reporting of sensitive behaviors among smartphone respondents...
  - ... Which we think may be related to enhanced data quality
  - ... And even surpassed that of R’s on computers who saw the standard implementation— the “benchmark”
... And why it leads us to a focus on questionnaire design

- Challenges of adapting a questionnaire between two survey systems—especially on a study with heavy smartphone use—brought to light design/layout decisions and constraints
  - Otherwise small effects may be magnified by small/touch screen

- While the software necessitated our switch to an “optimized” format, this was actually a questionnaire design issue that needs to be considered for this instrument—and others

- Layout-driven approach useful, especially here, given a long running questionnaire and limitation of the survey system...
  - ... but tangled with a lack of control— the “ebb and flow” of changing technology and ways in which survey software allows us to handle mobile respondents (or not)
Next steps: Stepping back/The bigger picture

- Currently using aspects of grids that we’ve identified as important to design experiments/develop models that take into account the “mobile space”
  - Crucial to improve upon this study’s limitations: Potential randomization of individuals to device, tighter control over formatting/layout, isolation of specific design elements, etc.

- However, implementation challenges we’ve described paint a complex (and accurate) picture of decisions faced on studies fielded with smartphone-heavy populations.

- Ultimate (and challenging!) goal: Designing clear, cogent questionnaires for all respondents—tailoring for, but not relying on, any one technology to achieve our goals of reliable and valid measurement.
References


