



Beyond Usability: Methodologies and Bias – Surveying the Surveys

Troy D. Abel^(✉)

University of North Texas, Denton, TX 76203, USA
troy.abel@unt.edu

Abstract. Surveys are a tool used in everyday usability testing to gather data and confirm findings amongst target populations; however, usability researchers must be aware of biases which can be introduced through certain types of survey questions. Survey instruments are a popular way to quickly gather qualitative and quantitative data from target populations for use in usability studies. These instruments may, at times, introduce unforeseen bias, and information participants may be unaware of when questions are triangulated and results combined during the analysis phase.

Certain populations may be hesitant to answer certain questions which may divulge immigrations status, sexual orientation, marriage status, age, and socioeconomic status to name a few. Of even greater concern occurs when answers to these, and other, questions are triangulated and combined to form inferences with regard to the population. For example, imagine the survey instrument contained questions regarding sexual orientation, marital status, and income levels – when triangulated and analyzed, the target population could be deemed marginalized and their results deemed unnecessary to the study, when in reality this population could potentially affect the outcome of the survey had their results been included.

Keywords: Usability · Bias · Marginalized populations · Idiosyncratic populations · Data triangulation · Usability testing bias · IRB

1 Introduction to Usability Testing Surveys and Tools

1.1 Usability Testing

According to Usability.gov, “Usability testing refers to evaluating a product or service by testing it with representative users. Typically, during a test, participants will try to complete typical tasks while observers watch, listen and takes notes. The goal is to identify any usability problems, collect qualitative and quantitative data and determine the participant’s satisfaction with the product” [1]. In addition to executing ‘task-based testing,’ quite often survey instruments are utilized to gather information about representative users. These surveys can take place prior to, or in tandem with, a task-based test. The results of these surveys are often analyzed in conjunction with the usability testing to uncover innate problems with a system.

1.2 The Survey

According to Merriam-Webster, a survey is “[meant] to query (someone) in order to collect data for the analysis of some aspect of a group or area” [2]. “A survey is a tool used to collect data on a large group of people to answer one or more research questions. It is one of the most common research tools used by social science research (sociology, political science, education, and psychology), health fields, and market research” [3]. Surveys are utilized to quickly gather data about target markets, delineate differences amongst populations, and specifically in usability testing, to understand unique differences with regard to usage and ultimate desires of a system.

Surveys are used for several reasons and a broad array of purposes. For example, surveys are utilized by the US Government for evaluating policies. “The Study of Income and Program Participation is a government-funded survey study designed to examine how well federal income support programs are operating. Survey questions from this study include information on housing subsidies, food subsidies, and income supports” [3]. Corporations use surveys to measure customer satisfaction. Social scientists use surveys for basic and applied research. Usability experts rely heavily on quantitative and qualitative data streams, particularly surveys to help define design requirements, drive product innovation, and test the usage of a given system. Usability surveys are mostly delivered via the internet (web and mobile), and several tools exist to help administer such surveys.

Usability studies often employ the use of surveys because they are versatile, cost effective, and generalizable. Surveys can be released and data obtained quickly, and cost-effectively, through the internet. Additionally, surveys are generalizable, “which means that the information found in the survey is representative or reflective of the entire population being studied—not just the sample collected,” [3]. if the proper sampling techniques are followed. Usability surveys often contain demographic information as well as socio-economic status information. In additional, several layers of additional self-identified data are collected. With all this data comes extreme ethical considerations to the researcher, the next topic of discussion.

1.3 Research Ethics in Corporate America

Ethics exist to protect the integrity of research, participants, and data obtained. Additionally, “... many of the ethical norms help to ensure that researchers can be held accountable to the public. For instance, federal policies on research misconduct, conflicts of interest, the human subjects’ protections, and animal care and use are necessary in order to make sure that researchers who are funded by public money can be held accountable to the public” [4]. Several different policy makers have their own set of Ethics standards for research, such as the National Science Foundation, National Institute of Health, and the Food and Drug Administration. All the entities have varying policies with regard to research and ethics. For purposes of this paper, “Research ethics are primarily concerned with three aspects of research: (1) the research participants themselves, (2) the data collected from the participants, and (3) how we present our findings” [3].

The research participants themselves should be given informed consent and not be harmed in any way, minimizing all possible risks to participants and maximizing benefits of participation. The data collected from the participants should be clearly documented, transparent, replicable, and stored securely. The findings should be responsibly interpreted and stick extremely close to the actual data.

Surveys collect large amounts of data, and often times, usability researchers are not trained in the skillful art of crafting survey questions, and therefore potentially bias their research. “Oftentimes, we think we know the answer to the research question, but we want to confirm it with evidence. Believe it or not, our personal beliefs and assumptions can influence the research design in such a way that the results will reflect our expectations. That is, we can accidentally bias our own research. In fact, even the research questions we ask can expose our biases” [3].

“In 1974, the National Research Act established the Institutional Review Board (IRB) system for regulating research with human beings, which exists today” [3]. The Institutional Review Boards (IRB) are in place to protect human subjects in applied research; however, quite often, usability researchers at private corporations don’t have an IRB to protect human subjects and approve their studies; therefore, these studies take place without regard for the participants. However, the IRB mandate only applied to federally funded research, and not private corporation research. Therefore, very few to no corporations, have IRB to review and approve human test subject research- this constitutes a major ethical flaw in Usability Experience (UX) research.

Often times, surveys ask sensitive questions without the necessary support from the sponsoring organization or pressure populations to participate unintentionally. For example, “Asking women questions about sexual abuse or rape might lead to serious emotional harm” [3]. Moreover, protected populations such as children, prisoners, LGBTQ+ participants may feel pressured to participate in surveys if asked.

Without IRB reviews and approvals of human test research, corporations are potentially biasing their data in more ways than one can imagine.

2 The Usability Survey

According to Perlman (2009), “Questionnaires have long been used to evaluate user interfaces (Root & Draper, 1983). Questionnaires have also long been used in electronic form (Perlman, 1985)” [5, 6]. These questionnaires have been used in usability to assist in uncovering problems innate to a designed system. According to Usability.gov, “When conducting an online survey, you have an opportunity to learn: (A) who your users are, (B) what your users want to accomplish, (C) what information your users are looking for” [7]. The structure of a usability survey is quite similar to that of a standard corporate survey: (1) demographics and market segmentation questions, (2) system satisfaction and investigation questions, and finally (3) follow-up questions which probe for deep data. Moreover, survey data, “... is often a main source of input for segmentation, personas, market feasibility, and decisions on prioritizing product functionality” [8].

3 Survey the Survey’s: Higher Education Surveys in Action

Currently, two (2) of the most popular surveys being utilized in Higher Education today are the National Survey of Student Engagement (NSSE) and The Freshman Survey (CIRP). These surveys are utilized to help administrators in higher education make changes to their campus in order to facilitate change based on survey responses. Both of these surveys demonstrate usage of potentially biasing questions. Two questions from each survey will be discussed below as examples of potentially biasing participants by reinforcing stereotypes (See Sect. 4.7).

3.1 The National Survey of Student Engagement (NSSE)

According to their website, “The NSSE survey, launched in 2000 and updated in 2013, assesses the extent to which students engage in educational practices associated with high levels of learning and development. The questionnaire collects information in five categories: (1) participation in dozens of educationally purposeful activities, (2) institutional requirements and the challenging nature of coursework, (3) perceptions of the college environment, (4) estimates of educational and personal growth since starting college, and (5) background and demographic information” [9]. This survey is unique in that it is extremely long and tedious for participants to fill out, and potentially suffers from significant participant drop-off.

The NSSE does not ask demographic information until the end of their long survey. After pages and pages of questions, participants are asked personally identifying information such as, “What is your gender identity?” The NSSE only provides 4 choices (See Fig. 1). One of these choices is “Another gender identity, please specify.” This is a potentially biasing question, and doesn’t explore all potential gender identities, possibly excluding some participants as even with an ‘other’ option some may feel ‘left-out.’



Fig. 1. Demographic question on the NSSE

A few questions below gender identity, the NSSE asks a participant to self-identify mental diagnoses such as sensory impairment, learning disabilities, or even mental health disorders (See Fig. 2). While the importance of this information can be debated, a UX researcher must ask themselves if it is important enough to potentially bias survey results.

Which of the following has been diagnosed? (Select all that apply.)

- A sensory impairment (vision or hearing)
- A mobility impairment
- A learning disability (e.g., ADHD, dyslexia)
- A mental health disorder
- A disability or impairment not listed above

Fig. 2. Demographic question on the NSSE

3.2 The Freshman Survey (TFS: CIRP)

The CIRP survey, according to their website, “... is designed for administration to incoming first-year students before they start classes at your institution. The instrument collects extensive information that allows for a snapshot of what your incoming students are like before they experience college” [10]. When exploring the questions, format, and flow of the CIRP, some participants may be immediately biased towards answering questions.

Take for example, question 1 and 2 of the CIRP. These questions ask participants to self-identify their gender and sexual orientation. Question 1 is: “What is your current gender identity?” (See Fig. 3) Beginning a survey about the new college experience with querying gender identity, may potentially bias some participants. Depending on how the survey is framed, where it is administered, and other factors, some participants may be ‘put-off’ by this as an initial question and think the entire survey will be about how they personally identify in the world. While this may be a question used to determine market segmentations and demographics, this research would argue placing this question later in the survey, or possibly at the end.

CIRP Freshman Survey

Sponsored by the Higher Education Research Institute and
Sample University

Percent completed

25 50 75

1. What is your current gender identity?

- Man
- Woman
- Trans man
- Trans woman
- Gender queer/Gender non-conforming
- Different identity

Next Page

Fig. 3. The first question on the web-based CIRP survey

This research notes, the CIRP takes into account best practices when listing multiple choice options with regard to gender and incorporates the most relevant terms in identification as well as a ‘Different identity’ choice. Question 2 of the CIRP is: “What is your sexual orientation?” (See Fig. 4) Again, this is a potentially biasing question to be asked as the second question of a new freshman in college, as they may perceive it as something major to do with the college experience.

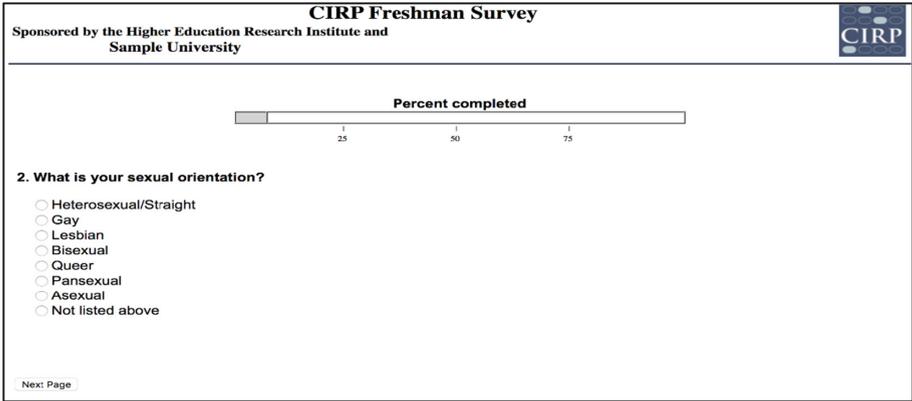


Fig. 4. The second question on the web-based CIRP survey

4 Analyzing Survey Data and Negotiating Bias

Jeff Sauro, an often-cited usability expert in usability statistics, has published what he calls the 9 biases that affect survey responses. These biases, "... can be particularly pernicious because they're harder to spot than more glaring problems In fact, there are not always clear remedies to the many biases that can affect your results. However, often just being aware of them is enough to help mitigate unwanted effects" [11]. Sauro lists these 9 survey biases: (1) Social Desirability & Conformity, (2) Yea Saying and Acquiescing, (3) Order Effects, (4) Prestige, (5) Threat & Hostility, (6) Sponsorship, (7) Stereotype, (8) Mindset (Carry-Over Effects), and (9) Motivated Forgetting [11]. Each of these will be discussed in detail below.

4.1 Survey Biases: Social Desirability and Conformity

Sauro describes this bias as one which specifically deals with social norms. Sauro says, "If it's socially acceptable ... respondents are much more likely to endorse and exaggerate. In addition to socially desirable, a number of studies show people will conform to group norms both offline and online. In fact, it's hard to convince respondents to go against what's acceptable even when things are clearly bizarre. This means respondents will have a propensity to provide the socially acceptable response over the true response" [11]. When the pressure of completing a survey weighs itself,

respondents find themselves subconsciously adhering to the social norms currently in existence.

When analyzing this type of bias, it's important to remember what social norms could influence a response for a particular question, and factor that information into the analysis of responses.

4.2 Survey Biases: Yea Saying and Acquiescing

This type of bias is characterized by those respondents who appear to over-answer 'yes' to a series of questions, even when the questions are negatively phrased and should normally elicit a 'no' response. Sauro says, "Respondents can tend to be agreeable (acquiesce) and respond usually positively to just about any question you ask them in a survey. One of the best way(s) to minimize this "yea" saying is to minimize simple yes-no answers and instead have respondents select from alternatives or use some type of force choice or ranking" [11]. Respondents tend to be amicable, and acquiesce, to satisfy what they believe others want to this.

When analyzing and uncovering this type of bias, it's often difficult to ascertain if a response is truly genuine or not.

4.3 Survey Biases: Order Effects

This bias specifically deals with the linear progression through a survey. Sauro states, "The order you ask questions matters. Mentioning products, brands, or events can affect how people rate their familiarity and attitudes on subsequent questions. This can be especially harmful in branding and awareness surveys as the mere exposure of a brand name first can influence later questions and findings. Response options also matter. A respondent might remember a choice that appeared in an earlier question and be more likely to select the response on later questions. You can often manage many order effects through properly sequenced questions and randomization" [11]. This level of bias is easily mitigated, by carefully scrutinizing the order and wording of usability survey questions.

4.4 Survey Biases: Prestige

Certain questions soliciting information with regard to socio-economic status (SES), income levels, and other 'prestigious' situations, should be carefully scrutinized. Sauro says, "Respondents will likely round up on income (especially men), education, and their reported power and prestige when making decisions. This is different than outright lying or cheating on a survey. If a question asks about prestige, assume the responses are inflated to present the respondent in a more favorable light. Exactly how much they are inflated will depend on the question, context and respondents" [11].

This type of bias, while at times easy to spot, is difficult to judge due to varying levels of respondent inflation and truthfulness.

4.5 Survey Biases: Threat and Hostility

Asking certain questions may trigger respondents and place them in a negative mindset for the rest of the survey. This bias is described by Sauro best, “Getting people to think about unpleasant things and events can get them in the wrong state of mind, which can cast a negative shadow on subsequent questions... Even rather benign questions (like asking people their marital status) may prime respondents with negative thoughts as participants recall bad past experiences (like a divorce or death in the family). Moving more sensitive demographic questions and anything that could potentially elicit negative thoughts to the end of a survey when possible may help” [11]. As Sauro suggests, placing these questions at the end of the survey may mitigate their effect on other answers.

4.6 Survey Biases: Sponsorship

If the survey’s sponsor is revealed before or during a survey, be prepared for automatic bias. Sauro says, “When respondents know where the survey is coming from (the sponsor), it will likely influence responses. [For example] if you know the questions about your online social media experience are coming from Facebook, your thoughts and feelings about Facebook will likely impact responses. This can be especially the case for more ethereal measures like brand attitude and awareness that can be affected from the mere reminder of a brand in the email invitation or name and logo on the welcome page. One of the best ways to minimize sponsorship bias is to obfuscate the sponsor as much as possible and/or use a third-party research firm...” [11]. The best way to negotiate this bias is to never reveal who the survey/research sponsor is to the participants.

4.7 Survey Biases: Stereotype

“Asking about gender, race, technical ability, education, or other socio-economic topics may reinforce stereotypes in the mind of the respondents and may even lead them to act in more stereotypical ways. For example, reminding people that stereotypes exist around those who are more technically averse (age), math ability (gender), or intelligence (education level) may affect later responses as the stereotype primes respondents through the questions” [11]. These types of questions potentially ‘put-off’ participants, and bias their responses throughout the survey. Questions which remind participants stereotypes exist should be avoided if possible, as these will likely bias survey responses for questions to follow.

4.8 Survey Biases: Mindset (Carry-Over Effects)

At times participants remember, or carry-over, remnants from previous questions or biases about a brand or situation they were asked about previously. This type of bias is known as Mindset according to Sauro [11]. To mitigate carry-over effects when designing survey instruments, consider spacing out questions or inserting ‘filler’ questions.

4.9 Survey Biases: Motivated Forgetting

“Memories are malleable and in general, we’re not terribly good at remembering events accurately. People tend to distort their memories to match current beliefs, also called telescoping. Respondents may recall an event but report that it happened earlier than it actually did (backward telescoping) or report that it happened more recently (forward telescoping). Many research questions rely on participants to recall specific events or behavior. There can be a tendency to recall events that didn’t happen or forget the specifics of an event” [11]. When asking participants to recall particular behaviors, be prepared for this type of bias: motivated forgetting. Participants will attempt to provide the ‘best answer’ to what they hope will be the response the survey is soliciting. Researchers should be aware of this type of bias when analyzing results.

4.10 Survey Inferences

Sauro says, “Just because a survey has bias doesn’t mean the results are meaningless”. Just because surveys have bias, doesn’t necessarily mean a researcher cannot use the resulting data. Sauro says, “Just because a survey has bias doesn’t mean the results are meaningless”. It does mean you should be able to understand how each may impact your results. This is especially important when you’re attempting to identify the percentage of a population (e.g. the actual percent that agree to statements, have certain demographics like higher income, or their actual influence on purchase decisions). While there’s not a magic cure for finding and removing all biases, being aware of them helps limit their negative impact. A future article will discuss some ideas for how to identify and reduce the effects of biases and other common pitfalls in survey design” [11]. Researchers should be aware of inferences and biases as to attempt to mitigate these types of biases.

5 Conclusion and Future Research

This research has demonstrated a need for corporate, academic, and private UX research teams to establish best practices with regard to Human Subject research. In addition, this research provides several classifications of bias, or inferences, in survey research and means to mitigate these biases. Future research includes creating a tool-kit for UX Researchers across all business types which will provide best-practices in UX survey creation, administration, and analysis, including human test subject’s ethics.

References

1. usability.gov Usability Testing. <https://www.usability.gov/how-to-and-tools/methods/usability-testing.html>. Accessed 20 Feb 2019
2. Merriam-Webster No Title. <https://www.merriam-webster.com/dictionary/survey>. Accessed 5 Feb 2019
3. Ruel, E.: 100 Questions (and Answers) About Survey Research. SAGE Publications, Inc., Thousand Oaks (2019)

4. NIH National Institute of Environmental Health Sciences. <https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm>. Accessed 15 Feb 2019
5. Perlman, G.: Electronic surveys. *Behav. Res. Methods Instrum. Comput.* **17**, 203–205 (1985). <https://doi.org/10.3758/BF03214383>
6. Root, R.W., Draper, S.: Questionnaires as a software evaluation tool. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI 1983*. ACM Press, New York, pp. 83–87 (1983)
7. Usability.gov Surveys. <http://usability.gov>. Accessed 16 Feb 2019
8. Sauro J Surveys. <https://measuringu.com/survey-biases/>. Accessed 18 Feb 2019
9. Indiana.edu NSSE. http://nsse.indiana.edu/html/survey_instruments.cfm. Accessed 1 Feb 2019
10. UCLA Higher Education Surveys: CIRP. <https://heri.ucla.edu/cirp-freshman-survey/>. Accessed 1 Feb 2019
11. Sauro J Survey Bias. <https://measuringu.com/survey-biases/>. Accessed 10 Feb 2019