Online Microsurveys for User Experience Research

Victoria Schwanda Sosik
Cornell University
301 College Ave.
Ithaca, NY 14850
vsosik@cs.cornell.edu

Elie Bursztein
Google Inc.
Mountain View, CA 94043
elieb@google.com

Sunny Consolvo
Google Inc.
Mountain View, CA 94043
sconsolvo@google.com

David Huffaker
Google Inc.
Mountain View, CA 94043
huffaker@google.com

Gueorgi Kossinets
Google Inc.
Mountain View, CA 94043
gkossinets@google.com

Kerwell Liao
Google Inc.
Mountain View, CA 94043
kerwell@google.com

Paul McDonald
Google Inc.
Mountain View, CA 94043
pmcdonald@google.com

Aaron Sedley
Google Inc.
Mountain View, CA 94043
asedley@google.com

Abstract

This case study presents a critical analysis of microsurveys as a method for conducting user experience research. We focus specifically on Google Consumer Surveys (GCS) and analyze a combination of log data and GCSs run by the authors to investigate how they are used, who the respondents are, and the quality of the data. We find that such microsurveys can be a great way to quickly and cheaply gather large amounts of survey data, but that there are pitfalls that user experience researchers should be aware of when using the method.

Author Keywords

Microsurveys; user experience research; user research methods

ACM Classification Keywords

H.5.2. User Interfaces: Theory and methods.

Introduction

To keep up with fast paced design and development teams, user researchers must develop a toolkit of methods to quickly and efficiently address research questions. One such method is the *microsurvey*, or a short survey of only one to three questions. There are several commercial microsurveys—including Google Consumer Surveys (GCS), SlimSurveys, and Survata—

Figure 1. An example of how a respondent encounters GCS. They are asked to answer a short survey question, or share the page they are reading via social media in order to continue reading the publisher's content.
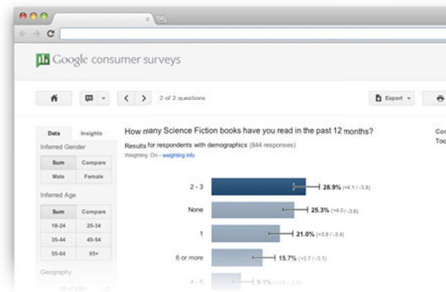

Figure 2. Part of the GCS results interface. To the left are controls to filter responses by demographics, and results to a multiple choice question are shown to the right.

that promise to provide people with large amounts of data quickly and at a relatively low cost. In this case study, we present a critical analysis of one type of microsurvey, Google Consumer Surveys, addressing questions about how they are being used, who their respondents are, and of what quality is the data they collect. We conclude with some current best practices for using this method in user research.

One Example of a Microsurvey: GCS

Since we use GCS in this case study, we first provide a brief overview of how it works. Each GCS respondent is shown only one question, two if there is a screening question. If a survey has more than one question, then each respondent is randomly shown only one of the survey questions. The survey designer can choose one of twelve predefined question formats that include open ended, single answer, multiple answer, and rating scale responses. Certain question formats allow for images in the question or responses. Questions and responses must be short, with 125 character and 44 character limits respectively; multiple choice questions are limited to showing 5 response options to each respondent.

Survey designers can request a representative population, or target respondents based on specific demographics (as inferred by IP addresses and DoubleClick cookies) or by using a screening question. Questions are then shown to people trying to access a publisher's premium content—primarily in the categories of News, Arts & Entertainment, and Reference—and people answer the question in order to continue reading the content (see Figure 1); in this way, these microsurveys are acting as a *surveywall* between the respondent and the content they want to access.

After data is collected, survey designers can view the results in the GCS interface, which provides users with basic analysis tools including comparison of results by different demographics and automatic, editable clustering of open-ended text responses (see Figure 2).

Results: Analysis of GCS

We analyzed GCS log data and data from several surveys run by the authors. Some of the surveys were run specifically to gather data about GCS as a method, and others were run to answer user research questions for our product teams, however we analyzed them from a methodological perspective for this case study.

GCS by the Numbers

GCS log data shows that the two most frequently used types of questions are multiple-choice questions (see Table 1). Together, single and multiple answers make up over 80% of all deployed GCS questions. However the most common question type—multiple answer—has the lowest completion rate (see Table 1).

On average, respondents spend 9.7 seconds responding to a GCS question, and the modal response time is 4 seconds (see Figure 3). GCSs also collect data very quickly—on average, surveys are approved to start collecting data between one and four hours after being created, and complete data collection in about two to four days. General population surveys finish data collection on the lower end of that range, whereas targeted surveys tend to take the four days.

Who are GCS Respondents?

In November 2012, PEW Research ran a study to compare GCS demographics with that of their

Demographic	PEW	GCS
Men	32%	27%
Women	35%	27%
18–24	33%	18%
25–34	37%	30%
35—44	49%	32%
45—54	38%	28%
55—64	28%	26%
65+	18%	23%
Unknown Age	—	27%

Table 2. Inferred GCS demographics compared to PEW demographics.

Survey Question	PEW	GCS
Do you ever use the internet to use a social networking site like MySpace, Facebook, or LinkedIn.com?	42% [age 50+]	46% [age 45+]
What is the primary social networking site you use?		Facebook (85%) LinkedIn (6%) Twitter (4%) Google+ (3%) MySpace (1%)

Table 3. Social Network usage among older Americans, using PEW and GCS survey samples.

Question Type	Usage	Completion Rate
Multiple answers	62.04%	20.56%
Single answer	21.71%	39.37%
Open Ended	4.62%	27.03%
Rating	3.81%	34.19%
Numeric open ended	1.60%	25.30%
Rating with text	1.50%	34.09%
Rating with image	1.30%	27.20%
Large image choice	0.99%	28.49%
Side-by-side images	0.92%	29.37%
Image with menu	0.82%	36.57%
Open ended with image	0.69%	27.79%
Two choices with image	--	--

Table 1. Rate of usage among survey designers and completion rate among respondents for the 12 different types of GCS questions.
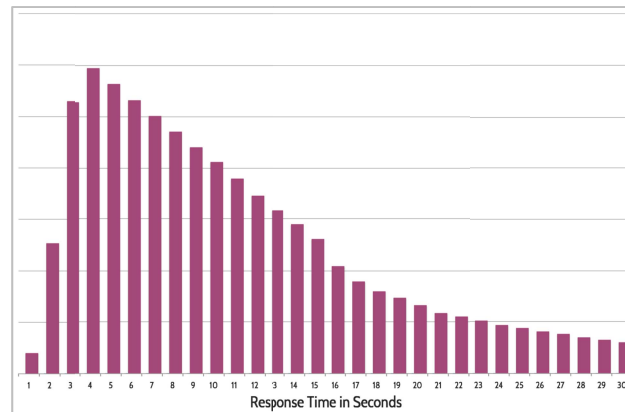

Figure 3. Distribution of response times in seconds to GCS survey questions.

telephone panels. Their overall findings were that GCS respondents "conform closely to the demographic composition of the overall internet population," and that there is little evidence that GCS is biased towards heavy internet users. [4]

We ran a series of GCSs to dig deeper into demographic and technology-use questions. We found that the rate of tablet ownership (PEW = 34%, GCS = 28%), cell phone ownership (91%, 67%) and use of cell phones (35%, 33%) or the internet for banking (61%, 48%) was lower among GCS than PEW respondents. In terms of demographics, GCS shows lower rates across age and gender (see Table 2). With respect to social networking site usage among older Americans, our findings using GCS were close to PEW (see Table 3).

We also compared GCS respondents to respondents from Survey Sampling International (SSI) and Knowledge Networks (KN) with respect to internet use and technology adoption. Results across the panels were similar, with SSI respondents tending to be the heavier internet users and technology adopters, and KN being the lowest (see Table 4).

Overall, while we notice demographic differences between the survey samples—likely due to the number of unknowns in GCS—technology usage and adoption is similar across all four samples, with PEW and KN representing the high and low extremes, respectively.

Respondents' Attitudes Toward Surveywalls
We ran a GCS to explore respondents' attitudes toward surveywalls that stand between them and content they are trying to access. We asked them which of five options they would prefer when trying to access premium content. We found that the most popular response was taking a short microsurvey (47%), followed by having content sponsored by an advertiser (34%), making a small one-time payment (10%), purchasing a subscription (6%), and other (3%; which they then had to specify as open ended text).

KN	GCS	SSI
For personal purposes, I normally use the Internet (5 = every hour or more, 1 = once per week or less)		
3.2	3.5	3.8
Other people often seek my ideas and advice regarding technology (5 = describes me very well, 1 = describes me very poorly)		
2.7	3.1	3.2
I am willing to pay more for the latest technology (same as above)		
2.3	2.6	3.1
Which of the following best describes when you buy or try out new technology? (5 = Among the first people, 1 = I am usually not interested)		
2.5	2.6	3.1
How frequently do you post on social networks? (5 = multiple times a day, 1 = once a month or less)		
1.7	2.1	2.4

Table 4. Technology use and adoption among 3 different survey panels.

Trap Questions in GCS

- What is the color of a red ball? (90.3% correct)
- What is the shape of a red ball? (85.7%)
- The purpose of this question is to assess your attentiveness to question wording. For this question please mark the 'Very Often' response. (72.5%)
- The purpose of this question is to assess your attentiveness to question wording. Ignore the question below, and select "blue" from the answers. What color is a basketball? (57%)

Data Quality: Survey Attentiveness

As one measure of data quality, we ran a GCS that asked respondents one of several trap questions. For a summary of how respondents performed, see the sidebar to the left. We find that our GCS respondents answered correctly the "Very Often" question less often (73%) than an example of the same trap question being asked on a paper survey (97%) [3]. A trap survey run in Mechanical Turk found only 61% of respondents answering correctly when asked to read an email and answer two questions [2], but this task is arguably harder than the questions we asked.

Data Quality: Garbage Open Ended Responses

We also analyzed data quality by looking at the rate of garbage responses that we received across 25 GCS questions run for other projects. Examples of these questions include: "which web browser(s) do you use?" and "what does clicking on this image allow you to do?" responses such as "blah", "who cares", and "zzzzz" and found that the percentage of garbage responses ranged from 1.8% to 23.4% (*Mean* = 7.8%). Our analysis revealed that the percentage of "I don't know" responses tended to correlate with the percentage of garbage responses, suggesting that people were more likely to provide such garbage responses when they were not sure of what the question was asking of them.

Conclusion: Best Practices for Microsurveys

We find that microsurveys such as Google Consumer Surveys can quickly provide large amounts of data with relatively low setup costs. We also see that the GCS population is fairly representative as compared to other large-scale survey panels.

However there are also pitfalls to keep in mind. Our findings from the trap question survey suggests that being concise is important to maximize data quality, which supports GCS's question length constraints. We also suggest that it is important to appropriately target surveys to a population in order to keep garbage open ended responses to a minimum. If respondents are being asked about something they are unfamiliar with, they are less likely to provide meaningful responses. Finally, multiple answer questions had the lowest completion rate—which is often used as a measure of data quality (e.g. [1])—so we suggest that people think critically about the types of questions they use, and consider using other question types if at all appropriate.

With respect to analyzing microsurveys, first it is important to remember that demographics are inferred, and there are many "unknowns". We also suggest using built-in text clustering tools to categorize open-ended responses, and if desired, following up with multiple choice questions to determine how frequent these categories are.

References
[1] Dillman, D. A. & Schaefer, D. R. (1998). Development of a standard e-mail methodology: results of an experiment. *Public Opinion Quarterly, 62*, 3.

[2] Downs, J.S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). Are your participants gaming the system? Screening mechanical turk workers. In *Proceedings of CHI '10*.

[3] Hargittai, E. (2005). Survey Measures of Web-Oriented Digital Literacy. *Social Science Computer Review. 23*, 3, 371–379.

[4] Pew Research (November 2012). *A Comparison of Results from Surveys by the Pew Research Center and Google Consumer Surveys.*