

Survey Research in HCI

Hendrik Müller, Aaron Sedley and Elizabeth Ferrall-Nunge

Neha Sharma and Mingkun Gao

Outline

- Background
 - History
 - When to Use
 - How to do Good Survey Research
 - Examples of Research
 - Summary
-

Background

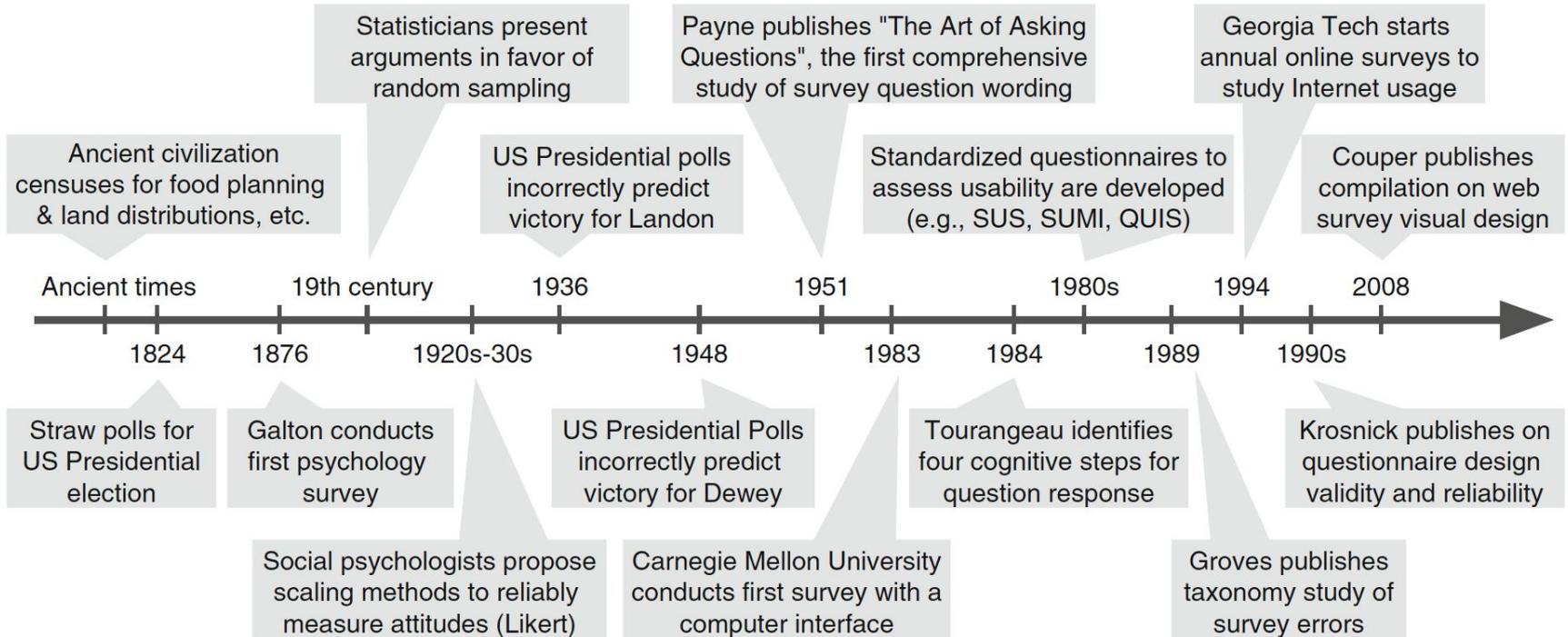


What is a survey?

- Gathering information
- From a subset of people
- Generalized to the target population



History of Using Survey



When to Use

- **Attitudes:** Measure and represent attitudes.
- **Intent:** Collect people's reasons for using an application.
- **Task success:** Quantify levels of success.
- **User experience feedback:** Collect open-ended feedback about a user's experience.

When to Use (Cont')

- **User characteristics:** Figure out users' characteristics.
- **Interactions with technology:** Understand how they interact with technologies.
- **Awareness:** Understand awareness of features.
- **Comparisons:** Compare attitudes across different applications.

When Avoid to Use

- Precise Behaviors
 - Use Log data
- Underlying motivations
 - Use contextual inquiry.
- Usability evaluations
 - Use Task-based observational research and interview methods, such as usability studies.

Using Survey with other Methods

- Follow previous qualitative studies to help quantify specific observations
- Initially identify high-level insights that can be followed by in-depth research through more qualitative methods

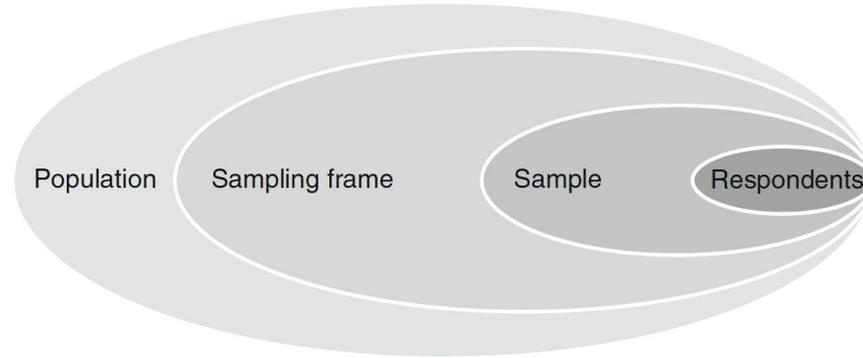
How to make a good survey?

- Research goals and constructs
 - Population and sampling
 - Questionnaire design and biases
 - Review and survey pretesting
 - Implementation and launch
 - Data analysis and reporting
- 

Research Goals and Constructs

- Define a clear research goal and constructs
 - Example of Research Goal and constructs
 - Research Goal: Understand users' happiness with an online application
 - Happiness is a multidimensional concept, separate it into measurable pieces (Constructs)
 - Constructs: overall satisfaction, perceived speed and perceived utility

Population and Sampling



- Population: People to whom researchers wish to generalize their results.
- Sampling frame: the set of people who the researcher is able to contact for the survey.
- Sample: the people from the sampling frame who are invited to take the survey.
- Respondents: samples who answer the survey.

Sampling Methods

- Sampling Methods
 - Probability (Random) Sampling
 - Gold standard because every person in the sampling frame has an equal chance
 - Avoid selection bias
 - Non-Probability Sampling
 - Impossible to reach and randomly select from the target population, e.g. small population, sensitive and rare behavior

Determining the Appropriate Sample Size

- Determine the size of population being studied.
- Choose the required level of precision.
 - Margin of error
 - Confidence level
 - E.g. In a study [CL = 95%, MoE=5%], 80% samples use computer everyday
- Estimate how many people to invite.
 - Response rate
- Invitation Methods: Mail, Phone, In-person and Internet.

Confidence level Size of population	90%				95%			
	10%	5%	3%	1%	10%	5%	3%	1%
10	9	10	10	10	9	10	10	10
100	41	73	88	99	49	80	92	99
1000	63	213	429	871	88	278	516	906
10,000	67	263	699	4035	95	370	964	4899
100,000	68	270	746	6335	96	383	1056	8762
1,000,000	68	270	751	6718	96	384	1066	9512
100,000,000	68	271	752	6763	96	384	1067	9594

Questionnaire Design and Biases

- Open-ended questions: Ask survey respondents to write in their own answers.
- Closed-ended questions: provide a set of predefined answers to choose from.

Questionnaire Bias

- Satisficing Bias: use a suboptimal amount of cognitive effort to answer questions.
- When this will happen:
 - Cognitive ability to answer is low
 - Motivation to answer is low
 - Question difficulty is high

(Satisficing)

To minimize Satisficing bias:

- Avoid too complex questions
 - Avoid options like “No Opinion”, “Don’t know” or “Unsure”
 - Avoid long questionnaire for increasing fatigue
 - Show importance of the survey to increase respondents motivation
 - Add some trap questions (e.g. Please input 5 in the following text box) to identify satisficers
-

Questionnaire Bias

- Acquiescence Bias: When presented with agree/disagree, yes/no, or true/false statements, some respondents are more likely to concur with the statement independent of its substance.
- When this will happen:
 - Agreeableness Personality
 - Social conventions, saying 'yes' is polite
 - Only think about why it is true
 - Assume the survey administrator agrees with the statement

(Acquiescence)

To minimize Acquiescence bias

- Avoid yes/no, agree/disagree or true/false questions
 - Ask construct-specific questions (i.e., questions that ask about the underlying construct in a neutral, non-leading way) instead of agreement statements
 - Use reverse-keyed constructs; i.e., the same construct is asked positively and negatively in the same survey. The raw scores of both responses are then combined to correct for acquiescence bias.
-

Questionnaire Bias

- Social Desirability Bias: respondents answer questions in a manner they feel will be positively perceived by others
- When this will happen:
 - Provide information for sensitive topics
 - Perceive a threat/consequence of answering truthfully
 - True identity was captured
 - The data is collected by another person directly

(Social Desirability)

To minimize Social Desirability bias

- Answer anonymously.
 - The survey should be self-administered.
-

Questionnaire Bias

- Response Order Bias: Select the items toward the beginning (i.e., primacy effect) or the end (i.e., recency effect) of an answer list or scale.
- Question Order Bias: the potential to bias subsequent questions by priming respondents.

(Order Related Bias)

To minimize Response Order Bias:

- Unrelated answer options should be randomly ordered across respondents

To minimize Question Order Bias:

- Questions should be ordered from broad to specific, similar to our conversation convention
 - Early questions should be easy to answer and directly related to the survey topic
 - The questionnaire should be divided into multiple pages with distinct sections labeled for easier cognitive processing
-

Other Types of Questions to Avoid

- Leading question
 - E.g. This application was recently ranked as number one in customer satisfaction. How satisfied are you with your experience today?
- Double-barreled question:
 - E.g. How satisfied are you with your smartphone and tablet computer?
- Recall question:
 - E.g. How many times did you use an Internet search engine over the past 6 months?
- Prediction question:
 - Over the next month, how frequently will you use an Internet search engine?

Visual Design Considerations

- Visual design of a survey, such as use of image, may affect participants' performance
 - E.g. Please rate your level of health. Image of patient vs. Image of Strong man
- Larger text fields - increased amount of information but may also intimidate respondents and can result in higher break-offs.

Leveraging Established Questionnaire

An alternative to constructing a brand new questionnaire is utilizing questionnaires developed by others.

Commonly used Questionnaires in HCI:

NASA Task Load Index (NASA TLX): Originally developed for aircraft cockpits, this questionnaire allows researchers to subjectively assess the workload of operators working with human-machine systems. It measures mental demand, physical demand, temporal demand, performance, effort, and frustration.

Launch Survey

- Review and Survey Pretesting: At this point in the survey life cycle, it is appropriate to have potential respondents take and evaluate the survey in order to identify any remaining points of confusion.
 - Cognitive testing
 - Field testing
- Implementation and Launch: Google surveys, survey monkey
 - How to maximize response rate: Make the survey as short as possible and provide incentives to participants.

Data Analysis and Report

- Data Preparation and Cleaning
 - Duplicate responses, Speeders, Straight-liners, Inadequate open-ended responses, handle missing data and break-offs.
- Analysis of Closed-Ended Responses
 - Descriptive statistics
 - Inferential statistics
 - Estimation statistics: margin of error, confidence interval.
 - Hypothesis testing.
- Analysis of Open-Ended Comments
 - Coding the responses: intra-coding reliability and inter-coding reliability.
- Report result
 - Answer the research question.
 - Describe original research goals and used survey methodology.

Examples of Research



“Why did you enroll in this course?” Developing a Standardized Survey Question for Reasons to Enroll

Developed for Massive Open Online Courses (MOOCs).

Survey research used to scale and personalize courses based on user’s motivations behind enrollment.

Responses were open-ended questions from previous courses and coded.

Table 1. Final ‘Why Enroll’ Survey Item

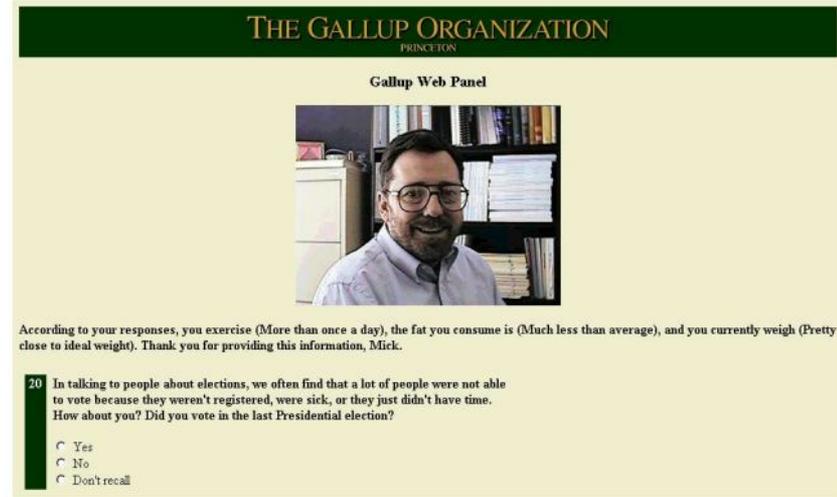
Why did you enroll in this course?	Applies	Does not apply
General interest in topic	<input type="radio"/>	<input type="radio"/>
Relevant to job	<input type="radio"/>	<input type="radio"/>
Relevant to school or degree program	<input type="radio"/>	<input type="radio"/>
Relevant to academic research	<input type="radio"/>	<input type="radio"/>
For personal growth and enrichment	<input type="radio"/>	<input type="radio"/>
For career change	<input type="radio"/>	<input type="radio"/>
For fun and challenge	<input type="radio"/>	<input type="radio"/>
Meet new people	<input type="radio"/>	<input type="radio"/>
Experience an online course	<input type="radio"/>	<input type="radio"/>
Earn a certificate/statement of accomplishment	<input type="radio"/>	<input type="radio"/>
Course offered by prestigious university/professor	<input type="radio"/>	<input type="radio"/>
Take with colleagues/friends	<input type="radio"/>	<input type="radio"/>
To improve my English skills	<input type="radio"/>	<input type="radio"/>

Social Presence in Web Surveys

Study to prove the merits of social interface theory on computer administered surveys (which are used to cancel social desirability effects).

Asked questions about socially desired traits and socially undesired traits.

Social interface effects could not be replicated.



The screenshot shows a survey interface for 'THE GALLUP ORGANIZATION PRINCETON'. At the top, it says 'Gallup Web Panel'. Below this is a video of a man with glasses and a light blue shirt, smiling. Under the video, there is a paragraph of text: 'According to your responses, you exercise (More than once a day), the fat you consume is (Much less than average), and you currently weigh (Pretty close to ideal weight). Thank you for providing this information, Mick.' Below this is a question: '20 In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, were sick, or they just didn't have time. How about you? Did you vote in the last Presidential election?'. The question has three radio button options: 'Yes', 'No', and 'Don't recall'.

Social Presence in Web Surveys (Cont'd)

Table 2. Scale Means by Condition (Standard Errors in Parentheses)

	Social Desirability	Impression Management	Gender Attitudes	Sensitive Admissions
Interaction	n.s.	n.s.	n.s.	n.s.
Low interaction	7.87 (0.14)	8.84 (0.19)	18.25 (0.16)	3.27 (0.07)
High interaction	7.83 (0.10)	8.91 (0.13)	17.98 (0.11)	3.30 (0.05)
Personalization	n.s.	n.s.	p<.05	n.s.
Logo	7.95 (0.10)	8.87 (0.13)	18.09 (0.12)	3.27 (0.05)
Male Picture	7.77 (0.09)	8.73 (0.13)	17.77 (0.11)	3.21 (0.05)
Female Picture	7.85 (0.09)	8.84 (0.13)	18.19 (0.11)	3.31 (0.05)

Table 3. Percentages on Behavior Variables by Condition

	% Used Cocaine in Lifetime	% Smoked Marijuana in Last Year	% Drink Daily or Almost Daily	% Attended Church Last Week	% Voted in Last Election
Interaction	n.s.	n.s.	n.s.	n.s.	n.s.
Low interaction	14.2	10.7	7.8	23.3	53.2
High interaction	15.3	10.2	7.7	25.7	52.2
Personalization	n.s.	n.s.	n.s.	n.s.	p<.05
Logo	15.4	10.8	7.4	23.2	52.8
Male Picture	14.7	9.9	8.0	24.3	55.3
Female Picture	14.2	10.5	7.7	26.1	49.7

Keep on Lockin' in the Free World: A Multi-National Comparison of Smartphone Locking

Participants in the survey answered three questions:

1. What secret unlock method do you use on your smartphone? [multiple choice]
2. Why did you decide (not) to protect your smartphone with a security mechanism (e.g., a 4-digit PIN)? [open-ended]
3. How sensitive is the information stored on your smartphone? [7-point scale, 7=very sensitive]

Research was conducted to understand the differences in attitude towards smartphone unlocking and indirectly information security.

Questions were translated in multiple languages.

Keep on Lockin' in the Free World: A Multi-National Comparison of Smartphone Locking

Variable	Estimate	Std. Err.	z	Odds Ratio
<i>Intercept</i>	.606	.054	11.18*	
country=AU	.326	.062	5.26*	1.39
country=CA	.270	.062	4.35*	1.31
country=DE	.409	.060	6.74*	1.51
country=IT	-.556	.057	-9.73*	0.57
country=JP	.067	.061	1.09	
country=NL	.306	.059	5.18*	1.36
country=UK	.565	.062	9.16*	1.76
group=SU	-.039	.032	-1.21	
Age=25-34	-.175	.042	-4.20*	.84
Age=35-44	-.159	.044	-3.66*	.85
Age=45-54	-.204	.051	-3.98*	.82
Age=55-64	-.407	.072	-5.66*	.66
Age=65+	-.530	.127	-4.17*	.59
Age=Unknown	-1.112	.659	-1.688	
Gender=Male	.323	.030	10.71*	1.38
Gender=Unknown	1.022	.461	2.21*	2.78

Table 1. Logistic regression model: likelihood of using a secure lock screen by country, age, and gender. An * denotes significance ($p < .05$).

Secure Unlock Responses		
1. Given protection goal	1,629	*
2. Protect against attacker (e. g. friends, children, thieves)	1,004	
3. Protect information	658	*
4. Protect from scenario	629	*
5. Protect certain action (e. g. calls, app use)	304	*
6. Lock is mandatory/recommended	105	*
7. Emotional reasons/sentiments	99	
8. Protection is necessary in general	47	*
9. Other reasons/don't know	215	
Slide-to-Unlock Responses		
1. Inconvenience	1,795	*
2. Absence of threat	1,340	*
3. Carelessness	381	*
4. Conflict with usage pattern	176	*
5. Protect using another measure	171	
6. Not secure anyway	90	*
7. Other reason/don't know	185	

Table 2. Major codebook categories and how many respondents mentioned them in total. An * denotes a significant omnibus difference in counts for each category between countries (Chi-square test, $p < .05$).

Amazon Mechanical Turk and Survey Research

“Thus, consistent with other research that has used MTurk, most workers are from the U.S. and India (Horton et al. 2011; Ipeirotis et al. 2010). While the studies examined here did not rule out other countries (e.g., Canada), the evidence at least suggests that requestors will be the most successful when recruiting workers from either India or the U.S.”

Table 3: Age, gender, and educational attainment levels

	U.S. Sample, Study 1	U.S. Sample, Study 4	India Sample, Study 5	Qualifying Survey	U.S. Population ¹
Sample Size	296	101	107	702	--
Time Period	August 2012	August 2013	August 2013	July 2013	2011/2012
Age					
18-29	41.55%	37.62%	41.12%	48.4%	22.0%
30-39	25.34%	30.69%	40.19%	25.7%	17.0%
40-49	16.89%	14.85%	8.41%	12.2%	18.2%
50-59	11.49%	9.90%	4.67%	10%	18.1%
60+	4.73%	6.93%	5.61%	3.6%	24.7%
Gender					
Male	42.6%	63.37%	64.49%	50.4%	49.1%
Female	57.1%	36.63%	35.51%	49.3%	50.9%
Education					
Some High School	N/A	0.99%	.93%	1.1%	8.58%
High School (or GED)	N/A	8.91%	.93%	11.1%	30.01%
Some College	N/A	32.67%	9.35%	29.4%	19.46%
College Graduate	N/A	47.52%	57.01%	48.8%	27.59%
Master's / Professional Degree	N/A	6.93%	31.78%	7.7%	8.4%
Doctorate	N/A	2.97%	--	1.9%	1.36%

Amazon Mechanical Turk and Survey Research (Cont'd)

Table 1: Qualifying survey HIT results

HIT #	Compensation	Time Available	Average Time Per Assignment	Completed
1	\$0.05	≈ 2 hours	1 minute, 29 seconds	10
2	\$0.11	≈ 1 day, 9 hours	2 minutes, 5 seconds	368
3	\$0.12	≈ 1 day	2 minutes, 9 seconds	106
4	\$0.13	≈ 19 hours	2 minutes, 25 seconds	200

“In July of 2013, a HIT for a survey that was similar in length and also limited to U.S. residents was created. Workers were paid \$0.50 for this HIT and it was completed in approximately five hours. The average time per assignment was nine minutes and 28 seconds. An identical HIT was created for residents of India with an average time per assignment of nine minutes and 27 seconds, virtually identical to the U.S. population.”

Summary



Make sure you have clear research goals and constructs, population and sampling, questionnaire design and biases, review and survey pretesting, implementation and launch, data analysis and reporting.

“Surveys are useful, but often misused by people who don't know the basics (where to show them, what to ask, who to ask, how to ask it). Often, people just want validation for a decision they've already made, or help making a decision they don't want to put real work into (experiments, qualitative user research).”

- Anonymous UX Researcher

References

Emily Schneider and René F. Kizilcec. 2014. "Why did you enroll in this course?": developing a standardized survey question for reasons to enroll. In *Proceedings of the first ACM conference on Learning @ scale conference (L@S '14)*. ACM, New York, NY, USA, 147-148.

DOI=<http://dx.doi.org.proxy2.library.illinois.edu/10.1145/2556325.2567852>

Marian Harbach, Alexander De Luca, Nathan Malkin, and Serge Egelman. 2016. Keep on Lockin' in the Free World: A Multi-National Comparison of Smartphone Locking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4823-4827. DOI:

<https://doi-org.proxy2.library.illinois.edu/10.1145/2858036.2858273>

Mick P. Couper, Roger Tourangeau, and Darby M. Steiger. 2001. Social presence in Web surveys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM, New York, NY, USA, 412-417. DOI=<http://dx.doi.org.proxy2.library.illinois.edu/10.1145/365024.365306>

References

Dupius, Marc, Barbara Endicott-Popovsky, and Robert Crossler. An Analysis Of The Use Of Amazon'S Mechanical Turk For Survey Research In The Cloud. University of Washington, 2017. Web. 1 Apr. 2017.
<http://faculty.washington.edu/marcjd/articles/An%20Analysis%20of%20the%20Use%20of%20Amazon%E2%80%99s%20Mechanical%20Turk%20for%20Survey%20Research%20in%20the%20Cloud%20-%20v.%203.pdf>