

Surveys

5

5.1 INTRODUCTION

Surveys are one of the most commonly used research methods, across all fields of research, not just human-computer interaction (HCI). Surveys are frequently used to describe populations, to explain behaviors, and to explore uncharted waters (Babbie, 1990). Surveys are also one of the most maligned methods. Surveys can be structured, well-tested, robust, and result in data with a high level of validity. However, surveys can be poorly done, resulting in data of questionable validity.

What is a survey? In short, it is a well-defined and well-written set of questions to which an individual is asked to respond. Surveys are typically self-administered by an individual, with no researcher present; because of this, the data collected is not as deep and in-depth as with other research methods (such as ethnography or focus groups). The strength of the survey is the ability to get a large number of responses quickly from a population of users that is geographically dispersed. Surveys allow you to capture the “big picture” relatively quickly, of how individuals are interacting with a certain technology, what problems they are facing, and what actions they are taking. Surveys also allow you to make statistically accurate estimates for a population, when structured using random sampling.

One of the reasons why surveys may be maligned is that they are often used not because they are the most appropriate method but because they are the easiest method. There are a lot of bad research projects, in which professors or students quickly write a survey, do not do sufficient pilot testing of the survey questions, distribute it to first-year students, and then claim that the survey results can generalize to other populations. Unless the actual focus of the research is university students, then this research example is misguided. As an example, an appropriate use of students was made in a survey study (Hanks et al., 2008), in which the goal of the research was to learn more about student perceptions of sustainable interaction design. It collected 435 surveys, from a cross-section of majors, not just computer science majors.

There are many HCI research projects in which a survey is the ideal method; in which the survey is well-designed, strict controls are used, and the resulting data has a high level of validity. Survey research may be the most appropriate methodology for measuring attitudes, awareness, intent, feedback about user experiences, characteristics of users, and over-time comparisons (Müller et al., 2014). Surveys may be less appropriate for precise measurements, or for solely identifying usability problems in an interface; however, surveys are often used appropriately in as one component of a

full evaluation involving user-based testing (described more in [Chapter 10](#)) ([Müller et al., 2014](#)). Since surveys primarily rely on users to self-administer, remember data that occurred in a previous point in time, and return the survey, without a researcher being physically present, there are a lot of background details that must receive attention for the data collected to be valid and useful.

Is a survey the same thing as a questionnaire? Well, many people do use the two terms interchangeably. Others differentiate between the “questionnaire,” which is the list of questions, and the “survey,” which is the complete methodological approach, including sampling, reminders, and incentives. For instance, Dillman states clearly that “the questionnaire is only one element of a well-done survey” ([Dillman, 2000](#), p. 149). While we acknowledge the difference, since the two terms are often used interchangeably, we use them interchangeably in this chapter.

5.2 BENEFITS AND DRAWBACKS OF SURVEYS

Surveys have many benefits and a few drawbacks. Using a survey, it is easy to collect data from a large number of people, at a relatively low cost. Surveys can be used for many different research goals. Because they allow access to a large number of people, surveys can be very useful for getting an overview, or a “snapshot,” of a user population. Surveys do not require advanced tools for development; they can be distributed easily using e-mail or existing survey websites, or done on paper. From a practical point of view, surveys are among the research methods most likely to get approval from an institutional review board or human subjects board because they are relatively unobtrusive (see [Chapter 15](#) for more information on institutional review boards).

There are a few drawbacks to using surveys as a research method. A survey is very good at getting limited “shallow” data from a large number of people but is not very good at getting “deep,” detailed data. Since surveys are typically self-administered (either on paper, e-mail, or websites), if interesting phenomena start appearing, it is usually not possible to ask follow-up questions, or go back and change the original survey instrument to ask more detailed questions.

Another major drawback is that surveys can sometimes lead to biased data when the questions are related to patterns of usage, rather than clear factual phenomena. For instance, a question such as the user's age or gender is not subject to interpretation or memory. Clearly, on a given day, an individual has an age (say, 33 years old) and a gender (male). However, questions related to mood (e.g., “How were you feeling when you were using this software application?”) are subject to recall bias if the event took place a significant amount of time earlier. Another example might be to ask people to recall how much money they have spent on e-commerce within a 6-month period or how many times they completed a certain task using a specific software application. Their response might be biased and either overestimate or underestimate the amount ([Andrews et al., 2003](#)). If data is of a factual nature and can instead be collected in an automated fashion using a computer, it may be a preferred

method compared to asking users to recall how many times they completed a task. In that type of situation, a combination of computer-collected data and a user survey might make the most sense (see the sidebar on photo tagging and sharing). It is also possible that the individuals who you are most interested in studying may come from a culture that is more oriented toward oral (spoken) approaches to communication than written approaches. If that is the case, then interviews or ethnography might be a more appropriate research method than a survey.

RESEARCHING PHOTO TAGGING AND SHARING BEHAVIORS

Two separate studies researching photo tagging and photo sharing behaviors illustrate how a combination of a survey and computer-collected data could be performed in a research study.

[Nov et al. \(2008\)](#) were interested in learning more about tagging behavior on Flickr (a website on which people can post pictures and notes [“tags”] about those pictures). The researchers were aware that it would not make sense to ask users how many tags they had created in a certain time period, as their responses were likely to be only a guess or an estimate, not an accurate count. However, there were a number of research questions that could best be investigated using a survey, so a combination of a survey and data logging, was used.

The researchers contacted a random sample of Flickr users who had posted at least five unique tags on pictures, in English (although this might have limited the sample to certain nationalities, the researchers wanted to make sure that the respondents understood the survey questions). A random sample of 1373 users was selected and e-mailed with an invitation to participate in a survey. At the end of the survey, respondents were asked to authorize the researchers to access data about tagging from their Flickr account (if the user gives permission, Flickr allows access to data from a user's account). Once the respondents filled out the survey and authorized access to their account data, the researchers were able to collect data on the number of tags. There were 237 valid survey responses and the average respondent had used 370 tags.

In a separate study related to Flickr usage published 8 years later, [Kairam et al. \(2016\)](#) were interested in studying how users chose to share photos. The researchers recruited participants who were already active Flickr users, to respond to a survey, requesting permission to access their photos as a part of the study (and the participants were paid a few dollars through Amazon's Mechanical Turk). There were 96 respondents to the survey. As a part of the research, 20 photos were selected from each participant's account, stratified across the five possible privacy settings for each picture. Of those 20 photos, 10 photos were randomly selected to be presented to the participants during the survey, with questions related to the privacy and content of the picture.

(Continued)

RESEARCHING PHOTO TAGGING AND SHARING BEHAVIORS—CONT'D

In data collection related to the research questions, but not related to the participants who responded to the previously described survey, more data was collected, from a sample of 638,930 active Flickr users (meaning at least one photo was uploaded in Jan. 2015), collecting aggregate data about their activity (e.g., number of photos, social connections, and group participation).

5.3 GOALS AND TARGETED USERS FOR SURVEY RESEARCH

Surveys are appropriate research methods for a number of different research goals. Since surveys are good for getting responses from large numbers of people, they are often used for collecting thousands, or even millions, of responses. The population of interest is also known as the “target population” (Couper, 2000) or, in the case of HCI research, the targeted users. If it is a well-defined population of interest, the actual number of individuals in the population can be identified. It might be a group of 20 individuals or 300 million individuals. However, if it is a well-defined population, there will be a definitive number of people within the population and it will be clear who is and who is not part of the population (Couper, 2000).

Who are the targeted respondents for the survey? Why are these people of interest? It is rare that you can truly say “anyone can respond to the survey.” Survey responses usually need to come from a specific population that is being studied—for instance, people of a certain age, users of a certain software application, people who work in a certain industry or company, or people who have a certain disability. You must first identify who they are and the limitations on this group. Do you limit responses to, say, people 30 years and older? People who are software engineers? People who have used the EndNote software application? People who are registered nurses? What demographic factors will decide whether a response from an individual is valid? It is important to note that the term “targeted respondents” from the world of survey design, can be used interchangeably with similar terms used throughout the book such as “user population” and “inclusion criteria.” The inclusion criteria will specify, in great detail, who qualifies to be included in your survey study (or in any other type of research study).

Once you have decided what criteria to use for your survey study, the next question is how can you find contact information for these individuals? Is there a well-defined list or directory of who these individuals are? General sociological research tends to use phone books or e-mail lists for the general public. When a listing or set of listings is used to define the potential survey respondents, this is known as “defining the population frame” (Couper, 2000). It is important to note that phone surveys, while they used to be more frequent, are now used much less often in survey research. There are several reasons for this: due to telemarketing calls, people do

not answer their phones as often; there are now several phones per individual; many people no longer have a landline phone; and government efforts in some countries have made it so that many individuals are placed on a “do not call” list because they do not want to receive many types of phone calls (Couper, 2005). When phone surveys are used now, there are often biases in the response.

For research into HCI, the population of interest is generally a bit more focused than just the general public, or a very broad set of criteria (e.g., registered voters). Often, there is a much more focused set of inclusion criteria. For example, if the inclusion criteria relate to being in a specific profession, websites, membership lists, and social networking for that profession, are great places to start. For instance, if the survey research targets researchers or practitioners in HCI, commonly used lists for HCI research are membership directories and social networking groups for professional organizations (such as SIGCHI, UXPA, and/or HFES). If inclusion criteria for a survey study relates to having a specific disability, membership directories of organizations for people with a specific impairment (such as organizations for people with spinal cord injuries) might be appropriate. If the survey relates to usage of a certain software application, lists of registered software users from a company might be appropriate. All of these types of lists may provide information on postal mailing addresses, phone numbers, or e-mail addresses. There may also be monthly or annual gatherings at which surveys, or information about surveys, can be distributed (Lazar, 2006). It is also possible that a website, online community, or social networking group might provide contact information for a group of potential respondents. Social networking applications can help recruit participants with a shared interest, for participation in a survey study (e.g., recruitment information can be posted on an interest group on Facebook or shared via someone on Twitter who has a lot of followers with a common interest). However, these methods alone may not work well for a lot of HCI research.

If the population for a survey is not easily well-defined, then the goal may be either to get a response that is diverse and represents multiple subgroups within the respondents or to get a survey response that matches what is known about the population (see Section 5.5).

5.4 PROBABILISTIC SAMPLING

The classic use of a survey in sociology is to make estimates for populations. The most accurate way to do this is by running a census, in which you attempt to get a survey response from every member of a population. Because a census is often very expensive and complex, they are not carried out very frequently. When a census is done, it tends to be sponsored by a large organization or governmental entity (see the US Census sidebar). If a population of interest is known and very small (say, up to a few thousand individuals), you might try to organize a modified census, in which everyone is invited to participate in the survey. However, it is not expected that everyone will take you up on the invitation and participate (Sue and Ritter, 2007).

US CENSUS—COUNTING EVERYONE

In the United States, a national census is taken every 10 years. Every person or family in the United States is supposed to fill out a paper survey. Responses to the Census Bureau are required by law, as the census count is used to distribute budgets and seats in congress and to make many governmental decisions. When a response is not received, individuals working for the Census Bureau visit residences to try and collect data from those who did not respond to the paper survey.

The Census Bureau tested a web-based form during the 2000 census. People who received the short form (five out of every six Americans) had the option of filling out the census form on the web. Each paper short form had an ID number. To ensure appropriate counting, the respondent had to enter the ID number on the web before filling out the actual survey.

Due to security and privacy concerns, the Census Bureau decided not to have a web-based form in 2010. However, the Census Bureau used a web survey for “reinterviewing” those who had already submitted their primary census form. See <https://www.census.gov/> for more information.

Instead of running a census, a structured method called “random sampling” (or “probability sampling”) is often used. In a probability sample, it is known exactly how likely it is for a participant to be selected for the sample, which is an equal, greater than zero chance, and everyone selected in the sample receives the same invitation to participate (Sue and Ritter, 2007; Müller et al., 2014). For instance, imagine that there are 10,000 members of a population of interest (the sampling frame). Perhaps 500 of these individuals are selected, at random, for requested inclusion in a survey study. All of these selected individuals must meet inclusion criteria (characteristics that they must have, such as being a nonsmoker or male) and not meet exclusion criteria (such as not being a native English speaker) (Sue and Ritter, 2007). See the sidebar on Random Sampling for an example of random sampling of a population of users.

A STUDY WITH RANDOM SAMPLING

When users are required to log into networked resources (such as an e-mail system, intranet, or social networking site), random sampling methods can be used, since a detailed list of who is considered to be within the population of interest does exist. For instance, a research study focused on Beehive, an enterprise social networking system from IBM. At the time of the study, it was estimated that there were at least 38,000 registered users of the site. A total of 500 users were randomly selected and invited to participate in the research study, based on having logged into Beehive during the last week and having enough data in their account so that friend recommendations could be made (the

inclusion criteria). Each selected user received a personalized survey, asking them to respond to recommendations made by the social networking software. During the period of the research study, 415 out of the 500 users logged in, 258 responded to the survey, and when the data was cleaned (due to incomplete or missing responses), 230 users had submitted valid surveys (Chen et al., 2009).

It is important to note that the sample frame need not be individuals; it can also be organizations. A long-term survey has documented the level of Internet access in public libraries across the United States. See the “Use of Sampling Frames in Studying Internet Access” sidebar for an example of a random sampling of organizations.

5.4.1 STRATIFICATION

Sometimes a sample can be stratified. A stratified sample is when you divide your entire population in separate subpopulations (known as strata) and a separate sample is selected from within each subpopulation (Frankel, 1983). So, when collected, data analysis can be made for each subpopulation and can be combined for the entire population. Stratification can help ensure that you have an appropriate number of responses from each subset of your user population (Babbie, 1990). Stratification can also help when the subpopulations have unequal representation, so that the relative distribution of the subpopulations in the sample can be increased or decreased (Frankel, 1983). A simple example is a random sample of university students. A random sample of university students is likely to have freshmen, sophomores, juniors, and seniors represented; however, due to enrollment trends often some class years have much larger classes of students as compared to other years, and therefore, there very likely would be an unequal number from each class who were randomly selected. However, a **stratified** random sample would have an equal number of respondents invited to participate, from each of those class years. A stratified random sample was used in a study of how people use technology to keep in touch after they move (see the Stratification sidebar for more information), so that long-distance moves would be represented more than local moves (which in fact are the majority of moves).

USE OF SAMPLING FRAMES IN STUDYING INTERNET ACCESS

The American Library Association sponsors a survey on the implementation and use of the Internet in public libraries in the United States. The earliest survey was in 1994 and the survey has been repeated on an annual or biennial basis since then. The most recent survey was in 2012, after which this survey was folded into the Digital Inclusion Survey (see <http://digitalinclusion.umd.edu/>). The survey started out as a paper-based survey but, over the years, it moved to a web-based survey.

(Continued)

USE OF SAMPLING FRAMES IN STUDYING INTERNET ACCESS—CONT'D

Since the survey is used to make national population estimates, the research approach used must be highly structured and controlled. Data was more recently collected using a web-based survey, with a paper letter mailed to public libraries to inform them about the existence of the survey. The letter included an identification code so that the survey data collected was identified to a specific library system or branch. The 2008 survey included 16,457 library outlets, a 6984 sample frame, and 5488 library responses (78.6%). The 2011–12 survey (with data published in 2012) included 16,776 library outlets, an 8790 sample frame (stratified by state and proportional by state and metropolitan status), and 7260 responses (82.5% response rate). See <http://www.plinternetsurvey.org/> for more information on methodological issues related to this survey.

STRATIFICATION

Shklovski et al. (2008) were interested in studying how technology influences the maintaining of friendships after a residential move. A sample of 6000 individuals was chosen from the US Postal Service's Change of Address Database. These were all individuals who had moved in the previous few months. The sample was stratified so that 1/3 of those selected had local moves, of 50 miles or less, while the other 2/3 selected had longer distance moves, of 50 miles or more. This stratification was done because the researchers were interested in studying long-distance moves. However, it is implied in their write-up that a majority of moves are local moves. Of the 6000 people selected from the database, 1779 (32%) responded to the survey. Two follow-up surveys were sent to the 1779 individuals who responded to the first survey. The second survey received 1156 responses, and a third survey received 910 responses. This research provides an example of stratification.

5.4.2 RESPONSE SIZE

If it is feasible for random sampling to be used in the research, this is preferable. However, the next question that comes up most often is, “how many responses do I need?” The statistics on this are not as clear as in, say, statistics in experimental design, where there is a clear threshold of significance or nonsignificance. In probabilistic sampling, the number of responses required depends on what level of confidence and margin of error are considered acceptable. For instance, for a simple random sample, a sample size of 384 may lead to a 95% confidence level with a $\pm 5\%$ margin of error (Sue and Ritter, 2007). That means that “if the survey were conducted 100 times, the true percentage would be within 5 percentage points of the sample percentage in about 95 of the 100 surveys” (Sue and Ritter, 2007, p. 30). To change the margin of error to $\pm 4\%$, 600 responses are needed; for $\pm 3\%$ margin of error, 1067

responses are needed. The margin of error is only valid using a true random sample. In this example, the actual size of the population sampled is irrelevant, since there is an automatic assumption that all populations being sampled are very large (Babbie, 2009). If the sample is relatively large compared to the population size (more than 5% or 10%), then the margin of error may be smaller, and can be calculated using the “finite population correction,” which is beyond the scope of this book. Another way to look at this is that, in a small population size, a smaller sample may be needed. See Sue and Ritter (2007), Babbie (2009), or Dillman (2000) for more information on appropriate sample sizes, confidence levels, and margins of error. The reader is especially encouraged to read (Müller et al., 2014, pp. 238–239), which is specifically focused on sample sizes in HCI research.

5.4.3 ERRORS

Random sampling seems like an ideal method but it is subject to a number of potential errors and biases. Careful attention to these potential problems can increase the accuracy and validity of the research findings. For instance, sampling error occurs when there are not enough responses from those surveyed to make accurate population estimates (e.g., if 10,000 individuals are surveyed but only 100 responses are received).

Coverage error occurs when not all members of the population of interest have an equal chance of being selected for the survey (e.g., if you use e-mail lists or phone lists to create the sample and not all potential respondents are on those e-mail or phone lists) (Couper, 2000). Measurement error occurs when survey questions are poorly worded or biased, leading to data of questionable quality.

Nonresponse error occurs when there are major differences (in demographics, such as age or gender) between the people who responded to a survey and the people who were sampled (e.g., if the sampling frame is split evenly by gender, but 90% of responses are from males) (Dillman, 2000).

5.5 NONPROBABILISTIC SAMPLING

The assumption in Section 5.4 on probabilistic sampling is that the goal is to achieve a population estimate. In HCI research, population estimates are generally not the goal. And so, users are more often recruited in a nonprobabilistic manner. Often, there is not a clear, well-defined population of potential respondents. There is not a list or a central repository of people who meet a certain qualification and could be respondents. For instance, due to requirements for patient confidentiality, it would be very hard to create a sample frame and a strict random sample involving people who have, for example, HIV (Müller et al., 2014). That may just be the nature of the population that no centralized list of potential respondents exists. So strict random sampling cannot be done. However, valid data can still be collected through nonprobability-based samples. Nonprobabilistic samples include approaches such as volunteer opt-in panels, self-selected surveys (where people often click on links on

a website or social networking), or snowball recruiting (where respondents recruit other potential respondents) (Müller et al., 2014).

It is important to note that different academic communities have different standards in how they apply sampling techniques. For instance, there are many people in the fields of social science and statistics who believe that without strict random sampling, no survey data is valid (Couper, 2000; Sue and Ritter, 2007). On the other hand, the HCI community has a long history of using surveys, in many different ways, without random sampling, and this is considered valid and acceptable. Part of this difference may stem from the nature of research in different communities. In some research communities, large national and international data sets are collected using rigorous, structured sampling methodologies. The general social survey in the United States (gss.norc.org) and the National Centre for Social Research in the United Kingdom (<http://www.natcen.ac.uk/>) are examples in the fields of sociology and public policy. Researchers can take these high-quality, probability-sampled data sets and perform analyses on the many variables in them. This is *not* the model of research used in HCI. In HCI, researchers must, typically, collect the data themselves. No large, well-structured data sets exist. The HCI researcher must go out, find users to take part in their research, and collect the data, as well as analyze the data. Because of this difference, both probability samples and nonprobability samples are considered valid in HCI research. There are a number of techniques for ensuring validity in nonprobability-based samples. The next sections detail the standard approaches for ensuring validity in nonprobability-based samples.

It is also important to note that, very often, surveys are used by HCI researchers, in conjunction with other research methods, when there is no claim of the representativeness of the survey responses, in fact, it is openly acknowledged that the responses represent a convenience sample. This is quite common, so, for instance, if you look at recent papers from the CHI conference, not only will you find surveys with over 1000 responses (such as Moser et al., 2016; Chilana et al., 2016), you will also find papers that combine small surveys with other research methods such as diary studies (Epstein et al., 2016), interviews (Dell and Kumar, 2016), usability testing (Kosmalla et al., 2016), and log analysis (Guy et al., 2016). These examples only scratch the surface; clearly, small, nonprobabilistic samples are used throughout HCI research on a regular basis, without concern.

5.5.1 DEMOGRAPHIC DATA

One way of determining the validity of survey responses is to ask respondents for a fair amount of demographic data. The goal should be to use the demographic data to ensure that either the responses represent a diverse, cross-section of respondents or the responses are somewhat representative of already-established, baseline data (if any exists). For instance, even basic demographic data on age, gender, education, job responsibility, or computer usage can help establish the validity and representativeness of survey responses when respondents are self-selected (Lazar and Preece, 2001). While this is not equivalent to the validity of a population estimate or random sampling, it is better than no check on the validity or representativeness of survey responses. Note

that, in some cases, researchers may have a goal to get representative data from multiple countries to do a multinational comparison. A great example of this is [Harbach et al.'s \(2016\)](#) recent study examining smartphone locking in eight different countries, with over 8000 survey responses. In such cases, it may be necessary to collect detailed demographic and cultural data related to the country, and researchers are also encouraged to consult a guide on doing cross-cultural HCI research (e.g., [Aykin, 2005](#)).

5.5.2 OVERSAMPLING

When there is not a well-defined list of users and strict random sampling is not possible, then the number of responses becomes increasingly important. For instance, in a nonprobabilistic sample, 20 survey responses may not be sufficient. Even with demographic data present, there may just be too many biases present, relating to which users have responded. However, when the survey response reaches a certain number that is considered large in proportion to the estimated or perceived population size, this can help establish some informal validity. This is known as *oversampling*. While not all researchers agree that oversampling increases validity ([Couper, 2000](#)), simply having a large response can reduce the likelihood of excluding any segment of the population ([Andrews et al., 2003](#)). However, the key is that the response must be large in the context of the population of interest. For instance, 500 survey responses would be a large number if the estimated total population of interest is around 5000 individuals. However, 500 survey responses would not be considered large if the population of interest is a country, such as Australia or France. One researcher suggests that 30 responses should be considered a baseline minimum number of responses for any type of survey research ([Sue and Ritter, 2007](#)). Fogg et al. used both demographic data and oversampling to learn more about web credibility in 2001 (see Demographic Data and Oversampling sidebar).

DEMOGRAPHIC DATA AND OVERSAMPLING

[Fogg et al. \(2001\)](#) wanted to learn more about how different elements of design on a website impact on the user's perception of credibility. To do this, they recruited survey responses through charitable groups in the United States and a news media organization in Finland. They received 1441 survey responses in 1 week. After discarding a number of responses due to inadequate information provided or responses that placed the respondent outside of the population frame, 1410 survey responses were considered valid.

The survey collected information on age, gender, country, education level, income, years on the Internet, average number of hours spent online per week, and average number of purchases online. The demographic information helped to confirm that the responses to the survey were, indeed, representative of the diversity of web users. The high number of responses helped to improve the validity of the study.

5.5.3 RANDOM SAMPLING OF USAGE, NOT USERS

Another approach to sampling is the random sampling of usage, not users (Lazar and Preece, 2001). For instance, it may be that every 10th time a web page is loaded, the user is asked to fill out a survey. Often, this survey appears in a pop-up window. This sampling technique is also known as intercept sampling (Sue and Ritter, 2007). While this gets an accurate picture of usage, a subset of users (those who use the web page often) is over-represented and those who do not view the web page often are under-represented.

5.5.4 SELF-SELECTED SURVEYS

In a “self-selected” survey, there is a link on a web page every time that it is loaded and everyone visiting the website is invited to fill out the survey. So, it is less about a certain group of people being recruited to participate and more about inviting everyone to participate. (Yes, it can be a bit fuzzy sometimes in nonprobabilistic surveys as to whether responses are invited or self-selected.)

If a self-selected survey is used, then both the number of survey responses and the demographic data on respondents become increasingly important in establishing the validity of the survey data. One of the earliest web-based survey studies came from the Georgia Institute of Technology. The entire population of web users was invited to participate. Banner ads about the survey, inviting people to participate, were placed on search engines, news sites, general advertising networks, mailing lists and newsgroups, and also in the popular press. Everyone was invited to participate in the surveys, which took place semiannually from 1994 to 1998. In the final survey, 5022 people responded. See Pitkow and Kehoe (1996) for a good summary of the studies and http://www.cc.gatech.edu/gvu/user_surveys/ for detailed data.

There may be reasons why a population could theoretically be well-defined and probabilistic sampling be used, but it is not logistically realistic. See the sidebar (“Probabilistic Sampling Probably Not Feasible”) for an example of a situation where self-selected surveys are the only feasible approach for a population that theoretically (but not realistically) could be sampled using probabilistic methods.

PROBABILISTIC SAMPLING PROBABLY NOT FEASIBLE

A well-known, ongoing survey in the accessibility community is the “Screen Reader Survey” run by WebAIM at Utah State University, which has the goal of learning more about the preferences of screen reader users. Screen readers are software applications, such as JAWS, VoiceOver, Window-Eyes, and NVDA, that allow people who are Blind or low vision to listen to the content on the screen, from web pages, applications, and operating systems (and they are not only installed on desktop and laptop computers, but also on tablet computers and smartphones). The first Screen Reader Survey was run in Dec. 2008 and Jan. 2009, with 1121 responses to the survey. The most recent (the 6th) Screen Reader Survey was run in Jul. 2015, with 2515 responses.

Theoretically, it might be possible to do a random sampling of current screen reader users (people who are blind or low vision and are already using a screen reader). However, to do so would require the major screen reader companies (companies such as Freedom Scientific, AI Squared, and Apple) to collaborate, share sales data, and any data that they have on users, to come up with a list of current screen reader users which probabilistic sampling methods could be applied to. Because the screen reader market is highly competitive, there have been lawsuits over intellectual property infringement, and there are a number of partnerships in place (e.g., AI Squared partnered with Microsoft to allow users of Microsoft Office to download free versions of the Window-Eyes screen reader), the likelihood of these companies collaborating on research and sharing data is not very high. Because companies want the Screen Reader Survey to report favorably on their market share, there can be pressure to “get out the vote.” While the Screen Reader Survey has a fair number of methodological flaws, it is the best set of data out there, over a 6-year period, about screen reader usage. For more information about the most recent Screen Reader Survey, see <http://webaim.org/projects/screenreadersurvey6/>. Chris Hofstader provides an in-depth criticism of the methodology of the screen reader survey at <http://chrishofstader.com/screen-reader-market-figures-my-analysis-of-webaim-survey-6/>. The inherent conflict between the need for the data and the wish for highly valid data when none is available can be seen in Chris’s comment that “I love numbers and, while the WebAIM survey has some major flaws, it is by far the best data we have available to us regarding the questions it covers” right before he provides pages and pages of criticism. ☺

Finally, it is important to note that self-selected, nonprobability-based surveys may be the most natural data collection method for investigating new user populations or new phenomena of usage. For instance, if no data exists about a certain user population or usage pattern, then a self-selected survey of users, asking about usage, might make the most sense, just as a starting point. The population of interest can be informed about the survey by posting a message about the survey to a social networking group, listserver, or chat room where members of the population are known to congregate (Schmidt, 1997).

5.5.5 UNINVESTIGATED POPULATIONS

Surprisingly, there are user groups that have still not been investigated in much detail. For instance, people with certain types of cognitive impairments have yet to receive much attention, if any, in the HCI research literature (see Chapter 16). For these populations where no baseline data exists, not enough is already known to develop hypotheses, experimental design, or well-structured time diaries. Population estimates may exist on how many people are living with a certain impairment within a certain country; however, no data exists on how many individuals with a certain

impairment are using computer technology. No baseline data exists, no estimate on the population size exists, no previous research exists, so all of the issues related to random sampling are really not appropriate. The goal of such a survey would be to establish the baseline data. In a case like this, the goal should be to simply get as large a response as possible. See the Computer Usage Patterns of People with Down Syndrome sidebar to see an example of where surveys were used to explore how young adults with Down Syndrome use computer technology.

COMPUTER USAGE PATTERNS OF PEOPLE WITH DOWN SYNDROME

When our team decided to pursue research about computer usage by people with Down syndrome in the mid-2000s, a search of multiple digital libraries and databases resulted in the determination that no research studies existed at that time, which examined how individuals with Down syndrome use computers and the Internet.

Only one design case study, where a website was being built to assist children with Down syndrome in learning about computers, was known to exist and be moving toward publication ([Kirijian et al., 2007](#)). Therefore, a survey methodology was the most appropriate approach, as a starting point for investigating the topic. We developed a large-scale survey, simply to gather baseline data about this user population. A 56-question survey was developed, covering demographic information, usage patterns, interaction techniques, and use of other electronic devices. Because it could be challenging to get accurate survey data from young adults with Down syndrome, it was decided that parents of children with Down syndrome would be recruited to respond to the survey.

The survey was placed on the web using survey monkey (a web-based tool), and responses were solicited through two organizations in the United States: the National Down Syndrome Congress and the National Down Syndrome Society. A total of 561 surveys were collected, which provides a rich foundation of data on which other studies and research projects can be built ([Feng et al., 2008](#)).

In communities where limited research has been done in the past, it may be challenging to find and recruit individuals to take part in the survey. There may be a lack of knowledge on the part of researchers, individuals may be reluctant to participate, or there might even be existing distrust.

Sometimes, snowball sampling can assist with getting survey responses. Snowball sampling is when individuals may not only respond to a survey, but also recruit someone else (usually a friend or colleague) to take part in the survey ([Sue and Ritter, 2007](#)). In a way, the role of contacting and recruiting participants shifts from the researchers to the survey respondents themselves. This method may work well when the population of interest is very small and hard to “break into,” and individuals in the population of interest may know each other well. An outside researcher, coming into a community of individuals, may not have a high level of credibility, but

another community member suggesting participation in a survey may come with a high level of credibility.

5.6 DEVELOPING SURVEY QUESTIONS

Once the goal and strategy for using a survey has been decided upon, the next step is to develop a survey tool. As mentioned earlier, some describe the survey tool itself as a “questionnaire.” The main challenge is to develop well-written, nonbiased questions. The questions in a survey can often lead to answers that do not represent what the researchers were actually asking. Since a majority of surveys are self-administered, they must be easy enough to understand that users can fill them out by themselves. In a limited number of situations, an interviewer may ask survey questions. For more information on interviews, read [Chapter 8](#).

It is important to understand that there are two different structures in a survey: the structure of single questions and the structure of the entire survey. More information on overall survey structure is presented in [Section 5.7](#). Most survey questions can be structured in one of three ways: as open-ended questions, closed-ended questions with ordered response categories, or closed-ended questions with unordered response categories ([Dillman, 2000](#)).

5.6.1 OPEN-ENDED QUESTIONS

Open-ended questions are useful in getting a better understanding of phenomena, because they give respondents complete flexibility in their answers. However, aside from the obvious drawback of more complex data analysis, open-ended questions must be carefully worded. Otherwise, they may lead to responses that either do not really help researchers address the root question, or responses that simply do not provide enough information. Consider the following open-ended question:

Why did you stop using the Banjee Software product?

This open-ended question provides no information about the possible causes; instead it requires the respondent to think deeply about what the causes might be ([Dillman, 2000](#)). The respondent may be too busy to come up with a complete response or may simply say something like “I didn't like the software.” It is a very broad question. More specific questions might be:

How did you feel about the usability (ease of use) of the Banjee software?

Did the Banjee software allow you to complete the tasks that you wanted to complete?

These questions address more specific topics: ease of use and task completion. The respondents cannot simply answer “I didn't like it,” although they could just answer “yes” or “no” to the second question. Perhaps another way to reword that second question might be as:

What barriers did you face, in attempting to use the Banjee software to complete your tasks?

In that revision, the respondents could simply say, “none” but the question also invites the respondents to think carefully about the problems that they might have faced.

5.6.2 CLOSED-ENDED QUESTIONS

There are two types of closed-ended questions. One type has ordered response categories, and the other type does not. An ordered response is when a number of choices can be given, which have some logical order (Dillman, 2000). For instance, using a scale such as “excellent to poor” or “strongly agree to strongly disagree” would be an ordered response. Likert scale questions, which often take the form of a scale of 1 to 5, 7, or 9, ask users to note where they fall on a scale of, for example, “strongly agree” to “strongly disagree.” Typically, closed-ended questions with an ordered response request respondents to choose only one item (see Figure 5.1).

Closed-ended questions with an unordered response allow for choices that do not have a logical order. For instance, asking about types of software applications, hardware items, user tasks, or even simple demographic information such as gender or type of Internet connection are unordered, but closed-ended questions. Figure 5.2 is an example of a closed-ended, unordered question.

What is your impression of using the website for www.veggieworld.com?

Please circle one number

Frustrating

Satisfying

1 2 3 4 5 6 7 8 9

FIGURE 5.1

A closed-ended question with an ordered response.

Source: QUIS, see <http://www.lap.umd.edu/quis/>.

Which application do you use most often for text editing? (please select only one)

- ☐ MS-Word
- ☐ WordPerfect
- ☐ Google Docs
- ☐ OpenOffice Writer
- ☐ WordPad
- ☐ QuickOffice
- ☐ Other (please specify)

FIGURE 5.2

A closed-ended question with an unordered response (single selection).

When using my primary computer, I use the following input devices or methods on a daily basis (select as many as apply)

☐ Keyboard

☐ Mouse

☐ Touchpad

☐ Trackball

☐ Voice recognition

☐ Multitouch screen

☐ Eye-tracking

FIGURE 5.3

A closed-ended question with an unordered response (multiple selection).

With unordered, closed-ended questions, you can often ask respondents to select more than one choice. On paper, this is not a challenge. However, it is important to note that if you are creating a web-based survey, different interface widgets must be used. Option buttons only allow one choice, whereas checkboxes allow for many choices. [Figure 5.3](#) is an example of a question that allows multiple responses.

5.6.3 COMMON PROBLEMS WITH SURVEY QUESTIONS

It is important to note that there are a number of common problems with survey questions. Researchers should carefully examine their questions to determine if any of these problems are present in their survey questions ([Babbie, 1990](#)):

- A “double-barreled question” asks two separate, and possibly related questions (e.g., “How long have you used the Word processing software and which advanced features have you used?”). These questions need to be separated.
- The use of negative words in questions (e.g., “Do you agree that the e-mail software is not easy to use?”) can cause confusion for the respondents.
- Biased wording in questions (such as starting a sentence with “Don’t you agree that ...”) can lead to biased responses. If a question begins by identifying the position of a well-respected person or organization (e.g., “Oprah Winfrey [or David Beckham] takes the view that ...”), this may also lead to a biased response.
- “Hot-button” words, such as “liberal,” “conservative,” “abortion,” and “terrorism,” can lead to biased responses.

5.7 OVERALL SURVEY STRUCTURE

Well-written questions are important, but so is the overall structure of the survey instrument. The questions do not exist in a vacuum, rather, they are part of an overall survey structure. For instance, a survey, in any format, must begin with instructions. These instructions must make clear how the respondent is to interact with the survey (Babbie, 1990). For instance, are respondents required to fill out all of the questions? Will respondents be required to enter their name or contact information? It is sometimes useful to put in a description, as a reminder, of who should be filling out the survey (e.g., you must be aged 65 years or older). If a survey is separated into multiple sections, then those divisions, and who should fill those different portions, must be made clear. Each section should be given an appropriate heading. Just as it is important to provide navigation on a website, a survey should provide navigation to the reader, whether the survey is paper, e-mail, or web-based. The user (respondent) needs to know where on the survey they should go, in what order. Sometimes, it is also helpful to provide contact information if the respondent has any questions (such as a telephone number or e-mail address).

Different formats of surveys (paper, e-mail, web-based) may require that information or instructions be presented to the respondent. For instance, in a paper survey, are there ovals or checkboxes? Should a checkmark be placed in them, should an X be placed in the box, or should the box be filled in? Should items be circled? Are respondents required to fill out all of the questions? These directions must be made clear. For an e-mail survey, should answers be typed in directly following the question on the same line, or on a line or two below it?

Layout of the survey instrument can also be important. For paper surveys, it is important to make sure that there is enough white space so that the respondent does not feel overwhelmed by the amount of information on a page (Babbie, 1990). Obviously, a balance needs to be struck. While respondents may worry if they see a 30-page survey, on the other hand, stuffing all of the survey questions onto two pages may prove to be problematic. Only white paper should be used, and a large enough font, in standard text, should be used (Dillman, 2000). Booklet printing (with two staples in the middle of the booklet) is preferred to one staple in the upper left-hand corner, but that is still preferred to any type of unusual folding or paper shapes that users may have trouble understanding (Dillman, 2000). In addition, do not use abbreviations to cut down on the amount of space needed, as they may cause confusion among respondents (Babbie, 1990). For a web-based survey, links are often provided directly in the survey, so that the respondent can click on the link and get a pop-up window with more detailed information. While pop-up windows are generally not good interface design, they work very well for giving short bits of information to users while they are in the process of responding to a survey.

Survey questions generally may be asked in any order which makes sense in the context of the research. However, it is important to keep in mind that questions relating to a similar topic or idea should be grouped together (Dillman, 2000). This tends to lower the cognitive load on respondents and allows them to think more deeply about the topic, rather than “switching gear” after every question. Because some

questions may require knowledge or details presented in other survey questions, it is generally hard to randomize the order of questions (Babbie, 1990). Rather, provide interesting questions at the beginning of the survey, to help motivate people to read the survey and complete it. Generally, it is a good idea to leave demographic questions until the end of the survey, as these are the least interesting (Babbie, 1990). Also, if there are any sensitive or potentially objectionable questions (relating to income, health, or similar topics), then they should be placed near the end, once the respondent has already become interested in the survey (Dillman, 2000). Note that survey length is an important consideration. While you want to include as many questions as possible on the survey, at some point, a survey becomes too long for many people to complete, and very long surveys can lead to very low response rates. Try to ask all of the questions that you need, but be reasonable when it comes to the amount of time that individuals need to set aside to respond to the survey.

The easiest type of survey is when all respondents should answer all questions. But frequently some questions do not apply to all respondents. For instance, imagine that you are running a survey to learn more about the e-mail usage habits of users over the age of 65. You may ask if they use a specific e-mail application (and you will need to be clear about the version of the application, and whether it is desktop, web-based, or smartphone-based). If the answer is “yes,” you may want them to answer a set of additional questions; if the answer is “no,” you want them to skip to the next set of questions. This is sometimes called a “contingent question” (Babbie, 1990) because the respondent’s need to respond to the second question is contingent on their response to the first question. This can be cause for confusion: if the directions and layout are not clear enough, a respondent who does not use Microsoft Office 365 may start reading questions relating to Microsoft Office 365 usage and be unsure of how to respond. On a paper survey, there are a number of ways to manage this. Babbie suggests using an indented box, with an arrow coming from the original question (see Figure 5.4). For a web-based survey, it may be possible either to provide a hyperlink to the next section (e.g., “If you answered no, please click here to move on to the next section”) or to automatically make a section of the survey “disappear,” so that the next question

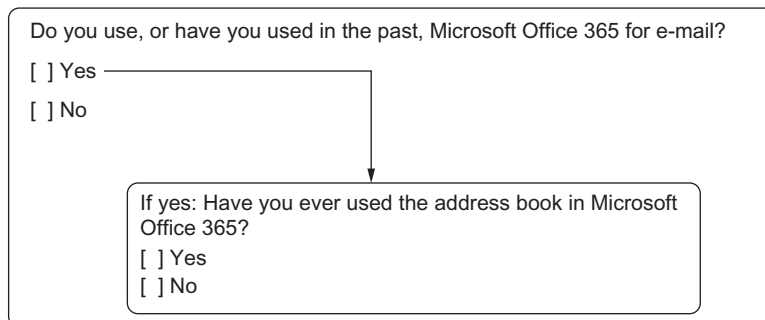


FIGURE 5.4

A contingent question on a paper survey.

presented is the one relevant to the respondent. This is similar to the “expand and collapse” menus that exist on many web pages. On a further note, the first question of the entire survey should always be a question that applies to everybody (Dillman, 2000).

5.8 EXISTING SURVEYS

It is important to note that there are many existing surveys that have already been tested and validated in the research literature in HCI. If a survey tool has already been developed, there is no need to create one from scratch.

For most research purposes, there will be a need to create a new survey tool. However, for tasks such as usability testing and evaluation, there are already a number of existing survey tools. Usually, these survey tools can be modified in minimal ways. For instance, one section of the survey tool can often be used independently of others. See Table 5.1 for a list of established survey tools.

Table 5.1 Survey Tools in HCI

Tool	Citations
Computer System Usability Questionnaire (CSUQ)	Lewis (1995)
Interface Consistency Testing Questionnaire (ICTQ)	Ozok and Salvendy (2001)
Perdue Usability Testing Questionnaire (PUTQ)	Lin et al. (1997)
Questionnaire for User Interaction Satisfaction (QUIS)	Chin et al. (1988) Slaughter et al. (1994) http://www.lap.umd.edu/quis/
Software Usability Measurement Inventory (SUMI)	http://sumi.uxp.ie/
Website Analysis and MeasureMent Inventory (WAMMI)	http://wammi.uxp.ie/

For more information about existing surveys for usability evaluation, the reader is encouraged to visit <http://garyperلمان.com/quest/>.

5.9 PAPER OR ONLINE SURVEYS?

An important question is to determine if you want to distribute surveys using paper, the web, e-mail, or a combination of the three. The traditional method is to use paper-based surveys. A benefit of this is that a majority of individuals can use a paper survey; however, people who are blind, visually impaired, or have a print-related disability will not be able to use a paper survey (see Chapter 16 for more information on doing research with computer users with disabilities). If you only use an electronic survey (web or e-mail), you are automatically cutting out any potential respondents who do not have access to a computer and a network, which may include users who are economically disadvantaged, or ethnic or racial groups that have lower base rates of computer access (Andrews et al., 2003). In addition, if you are creating an electronic survey, you must make sure that the interface is usable by a wide range of individuals who may respond to your survey (such as users with disabilities and older users).

In reality, the relative strengths and weaknesses of online and paper surveys generally do not influence which one is used. One major influence on which method (or combination) is used is how the researchers have best access to the user population of interest. In some cases, the best access is to visit individuals at a weekly meeting where paper surveys can be passed out. In other situations, if a list of postal mailing addresses exists for potential respondents, paper surveys can be mailed. If a list of e-mail addresses exists, e-mailed surveys may be best. Another major influence on which method (paper, e-mail, or web) is used, is the ease of developing online surveys, using existing web-based tools which allow a survey to be posted and distributed in minutes (although, honestly, such a rushed job is likely to lead to design problems and errors). These web-based tools include SurveyMonkey, SurveyPlanet, FreeOnlineForms.com, and Google Forms, and they certainly influence many researchers to use a web-based survey, even if there are good methodological reasons to choose another format. Web-based surveys are now the most frequently used format of survey research.

Sometimes, a combination of paper and web-based surveys can be used to make sure that all portions of a target population are reached (Lazar et al., 1999). It is also sometimes helpful to offer respondents a choice between a paper and an electronic version of the survey, as some research suggests that some people may simply prefer filling out surveys on paper (Schonlau et al., 2003). These mixed-model designs, in which paper, e-mail, and web-based versions of a survey instrument are used together, can help improve the response rate, but caution must be taken to make sure that no biases are introduced into the data collection process (from three survey instruments that, in fact, do have minor differences) (Couper, 2005). Obviously, paper surveys must be used to study questions such as “why don’t people go online?” and other research questions related to nonuse of technology (Lazar and Preece, 2001). Another potential complication is that you may need to offer your survey in multiple languages. In countries where there are multiple official languages, this may be a legal requirement (Normand et al., 2014). In other cases, you may be interested in studying a group of computer users who do not share the same primary language. If so, you need to ensure that the surveys in two or three different languages are in fact asking the same questions and that there are no mistranslations. Professional human translation is necessary in such a scenario (automated tools for translation are not sufficient for the task).

There are benefits to electronic (both e-mail and web-based) surveys. Copying costs, mailing, and related postage costs can be eliminated with electronic surveys (perhaps having only the cost of sending out paper letters notifying potential participants, when needed). While the set-up costs may be high for a custom developed web-based survey, using existing web-based survey tools when possible, make web-based surveys the most cost effective in terms of time and expenses (Sue and Ritter, 2007). In most cases, web-based surveys and even e-mailed surveys can automatically have responses saved in a spreadsheet or database, eliminating the need for time-consuming data entry and eliminating many data entry errors (Lazar and Preece, 2001). While response rates in online surveys may sometime be lower, the speed of response is certainly higher (Sue and Ritter, 2007), as is the speed of analysis by researchers (Müller et al., 2014).

The question is often asked if the responses from electronic (web-based or e-mail) surveys are as trustworthy or valid as paper surveys. There is no evidence to suggest that people are more dishonest in online surveys than in paper surveys, as people can lie easily in both. However, there is evidence that people, when delivering bad news, are more honest in online communication than face to face ([Sussman and Sproull, 1999](#)). There is also evidence that people, when they care about a topic, are likely to be very honest. If the surveys can be submitted anonymously, this may also lead to an increased level of self-disclosure ([McKenna and Bargh, 2000](#); [Spears and Lea, 1994](#)). Therefore, web-based surveys can sometimes be superior to e-mailed surveys (which clearly identify the respondent) for dealing with sensitive information ([Sue and Ritter, 2007](#)). In addition, respondents to self-administered surveys tend to provide more honest answers to sensitive questions than in interviews ([Couper, 2005](#)). Overall, the likelihood that someone will lie in an electronic survey is the same as the likelihood that someone will lie in a paper-based survey.

In traditional paper-based surveys, individuals may have to sign an “informed consent form” (also known as an institutional review board [IRB] or human subjects form), acknowledging that they are aware that they are taking part in a research project and giving their consent. There is debate as to how individuals can best give informed consent when they respond to a survey online. For more information on informed consent online, please see [Chapter 15](#).

5.10 PILOT TESTING THE SURVEY TOOL

After a survey tool is developed, it is very important to do a pilot study (also known as pretesting the survey) to help ensure that the questions are clear and unambiguous. There are really two different areas of interest within a pilot study: the questions themselves and the interface of the survey. While the interface features primarily refer to web-based or e-mailed surveys, there are also interface features on paper-based surveys. For instance, on a paper survey, there should be an examination of issues such as the font face and type size, spacing, use of grids, and cover designs ([Dillman, 2000](#)). While these are theoretically different pilot testing sessions for the questions and for the layout, in reality, they take place at the same time. See [Chapter 10](#) for more information on usability testing of a computer interface.

[Dillman \(2000\)](#) suggests a three-stage process of pretesting a survey, while noting that it is rarely done thoroughly. The three stages are as follows:

1. Review of the survey tool by knowledgeable colleagues and analysts.
2. Interviews with potential respondents to evaluate cognitive and motivational qualities in the survey tool.
3. Pilot study of both the survey tool and implementation procedures.

The idea of this three-stage process is that you start first with people who are knowledgeable, but are not potential respondents. (Note that you start first with expert nonrespondents, just as in usability testing in [Chapter 10](#).) You begin with

expert evaluations before involving any representative users. You then ask a few potential respondents about the clarity and motivation of the questions in the survey. Finally, you do a pilot study where potential respondents complete an entire survey and the researchers can note any flaws. While this three-stage process is ideal, in reality, most research in HCI involves either a few colleagues examining the survey tool or a few potential respondents reading over the survey tool and giving some feedback, but even at this minimal level, the pilot study is still necessary and important.

A pilot study can help the researcher identify questions that are confusing or misleading. These pilot study efforts are aimed at determining the validity of the survey, that is, does the survey measure what it is claiming to measure? (Babbie, 1990; Ozok, 2007). There are usually a few common problems discovered in a pilot study, to keep an eye out for. For instance, questions that were not answered, questions where multiple answers were given (when only one was expected); and questions where respondents filled out “other” (Babbie, 1990). All of these are signs that a question might need to be reworded. A pilot study, ideally, will involve a small number of potential respondents (people who meet the inclusion criteria) answering the survey questions, with encouragement to provide specific feedback on the questions in the survey. For a small survey study (say, where the goal is 200–300 responses), perhaps 5–10 people taking part in the pilot study would be sufficient. However, for larger survey studies, where the goal is 100,000 survey responses, a corresponding larger number of individuals should take part in the pilot study. It is important to note that individuals who responded to the pilot study should generally not take part in the main study and their data should not be included. The process of participating in the pilot study could bias the future responses and therefore, they should not be included in the main data collection.

A different type of evaluation can take place at a later time. When a survey instrument has been used to collect data multiple times, then the reliability of that survey can be established. Reliability is the determination of whether a survey measures constructs consistently across time (Babbie, 1990; Ozok, 2007). Methods for measuring the internal reliability of questions, such as having the same question asked multiple times in a different way, can be used. The Cronbach's alpha coefficient is often used in that situation (Ozok, 2007).

Another approach to evaluating survey questions after data is collected from many people, especially if the survey has a large number of questions, is *exploratory factor analysis*. In factor analysis, statistical software creates an artificial dimension that would correlate highly with a set of chosen survey question data (Babbie, 1990). Researchers then determine how important the specific survey question is, based on the factor loading, which is the correlation level between the data item and the artificial dimension. Survey items with high factor loadings have high correlation, and are likely to be more predictive, and therefore, more relevant (Babbie, 1990). Exploratory factor analysis can help to cut down the number of questions in a survey (Ozok and Salvendy, 2001). For instance, in one of the two research projects described in the Flickr sidebar, the survey questions were validated using an

exploratory factor analysis of 193 users. “Items showing factor loading higher than 0.6 and cross-loadings lower than 0.4 were retained, and others were dropped” (Nov et al., 2008, p. 1098).

5.11 RESPONSE RATE

A good sampling method and a well-written survey tool are important. However, those steps alone do not guarantee a sufficient number of responses to a survey. One of the main challenges of survey research is how to ensure a sufficient response rate. Other research methods tend to have fewer users taking part and higher incentives for taking part, than in survey research. For instance, if 70 people take part in an experimental research study, they may each be paid \$100 for their participation. Obviously, this is not feasible when thousands of individuals are responding to a survey. Perhaps to increase the response rate, the names of respondents could be entered into a drawing to win a prize. Also, surveys are generally self-administered, regardless of whether they are paper, e-mail or web-based. Individuals often need to remember where the survey is located (the URL for a web-based survey, or where they have put the paper survey) and complete it in a timely manner, with the caveat being that they may not receive any major incentive for doing so. So it is important to motivate people to respond to surveys.

There are a number of tried and tested ways to increase the response rate to a survey. For all types of survey (paper, e-mail, and web-based), there should be some type of introductory letter, letting individuals know that they have been selected for inclusion in a survey study. The letter should tell people: who is sponsoring the research study, why it is important, what the expected timeframe is, and hopefully establish some authority or credibility. This is not the same thing as an informed consent form, this is all about establishing the importance and credibility of the survey study, to motivate people to respond. For instance, if an individual who is a trusted authority within the community of individuals helps to introduce the survey, this may help increase the response rate. Or if the survey comes from a well-respected government source, this should be clearly identified to help establish authority.

Aside from establishing the credibility of a survey, another method for increasing the response rate is to increase the ease in returning a survey. For instance, a paper survey should be accompanied by a self-addressed return envelope with postage included.

A multistep contact process tends to increase the response rate. Researchers should make multiple contacts with respondents. For instance, Dillman (2000) suggests the following process for paper surveys:

1. Send a precontact letter (usually with information from a trusted authority, as stated earlier), before the actual mailing.
2. Send a postal mailing, which includes the actual survey.
3. Send a thank you postcard (which thanks people for their time and serves as a reminder).

4. Send a replacement survey to nonrespondents 2–4 weeks after the original one was sent.
5. Make a final contact using a different mode. If the original survey was sent using postal mail, then maybe a phone call or e-mail should be used. If the survey was electronic, maybe a postal letter or phone call should be used. The idea is to have a different delivery method for the final contact that gets the attention of the respondent.

Depending on how the researchers have access to the potential respondents, different methods of postal mail, e-mail, phone calls, or even instant messaging, may be interchanged. So, for instance, in an electronic survey (web-based or e-mail), multiple reminders are important, and each time, the researchers should give the survey instrument (for a web-based survey).

A common question, mentioned earlier in the chapter, is the question “How many survey responses are enough?” This is not easy to answer, as it has to do with a number of different issues: What is the goal of the survey? What type of survey? What sampling method has been used? What level of confidence and margin of error is considered acceptable? See [Section 5.4.2](#) earlier in this chapter, where these questions are discussed.

5.12 DATA ANALYSIS

There are several ways to analyze survey data. The analysis chosen will depend, in large part, on:

- whether it was a probabilistic or nonprobabilistic survey;
- how many responses were received; and
- whether a majority of questions were open-ended or closed-ended questions.

Generally, the quantitative and qualitative data is separated for analysis. The data is “cleaned,” meaning that the researchers look through and make sure that each survey response is valid, and that none of the responses are either repeats (where the same person submitted more than one response), incomplete (where most questions were not answered), or invalid (due to a respondent not meeting the qualifications). The quantitative data is ready to analyze, whereas the qualitative data must first be coded (see [Chapter 11](#) for more information on content analysis).

Often, the goal of quantitative data analysis is simply to have a set of “descriptive statistics” that simply describe the data collected in a manageable way ([Babbie, 1990](#)). No one but the researchers will read through every survey response so the descriptive statistics are simply a short, high-level summary of the data. Most often, descriptive statistics involve percentages, ratios, or matrices. Inferential statistics involve a higher level of understanding of the data, by understanding the relationships between variables and how they impact each other. For more information on statistical analysis, read [Chapter 4](#).

5.13 SUMMARY

Surveys are a very powerful tool for collecting data from many individuals; however, there are only certain types of research questions in HCI for which surveys are the most appropriate research method. For an appropriate survey method, there must be a number of different steps that take place. To ensure validity and reliability, survey questions must be pilot tested, to ensure that they are clear, unambiguous, and unbiased. The overall survey design should make it easy for respondents to understand and use the instrument, whether web-based, e-mailed, or on paper. Appropriate sampling methods, even if they are nonprobabilistic, must be used to ensure a representative response that can answer the research questions. Good introductions, establishing the credibility and importance of the survey, as well as providing ongoing reminders to respond, can increase the likelihood that there will be a sufficient number of responses for the data to be considered valid. Other research methods can also be useful in conjunction with surveys, such as focus groups, interviews, or time diaries.

DISCUSSION QUESTIONS

1. Is a survey the same thing as a questionnaire? If not, how are they different?
2. What is the difference between the target population and the sampling frame?
3. Why are censuses done rarely? What is often used instead when population estimates need to be made?
4. What is the defining characteristic of a probability sample?
5. What is a stratified random sample? How is it different from a traditional random sample?
6. What is one of the major reasons that nonprobabilistic sampling is considered appropriate in human-computer interaction research but not in other research communities?
7. What is oversampling and why might it help improve validity of the research?
8. What is the difference between an open-ended and a closed-ended question?
9. Why might you want to use an existing survey instrument, when possible?
10. What is a double-barreled question and why is it not a good idea?
11. What is a contingent question and how might you deal with one in a survey layout?
12. What are two methods for testing a survey tool?

RESEARCH DESIGN EXERCISE

Consider that you want to learn more about how people use USB portable storage, as compared to network/cloud-based storage (in web accounts) or storage on their computer hard drives. More specifically, you want to learn how children and young adults (aged 10–18) and older users (aged 65–85) use these storage devices. You want to learn more about which devices individuals prefer and in what situations they use them. Would you use probabilistic or nonprobabilistic sampling? What questions might you ask? Come up with at least five questions. How would you structure the survey? Would you use contingent questions? How would you pretest the survey? How would you ensure that you receive a sufficient number of responses?

REFERENCES

- Andrews, D., Nonnecke, B., Preece, J., 2003. Electronic survey methodology: a case study in reaching hard to involve Internet users. *International Journal of Human-Computer Interaction* 16 (2), 185–210.
- Aykin, N. (Ed.), 2005. *Usability and Internationalization of Information Technology*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Babbie, E., 1990. *Survey Research Methods*, second ed. Wadsworth Publishing, Belmont, CA.
- Babbie, E., 2009. *The Practice of Social Research*. Wadsworth Publishing, Belmont, CA.
- Chen, C., Geyer, W., Dugan, C., Muller, M., Guy, I., 2009. Make new friends, but keep the old: recommending people on social networking sites. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 201–210.
- Chilana, P., Singh, R., Guo, P., 2016. Understanding conversational programmers: a perspective from the software industry. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, pp. 1462–1472.
- Chin, J.P., Diehl, V.A., Norman, K.L., 1988. Development of an instrument measuring user satisfaction of the human-computer interface. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 213–218.
- Couper, M., 2000. Web-based surveys: a review of issues and approaches. *Public Opinion Quarterly* 64, 464–494.
- Couper, M., 2005. Technology trends in survey data collection. *Social Science Computer Review* 23 (4), 486–501.
- Dell, N., Kumar, N., 2016. The ins and outs of HCI for development. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, pp. 2220–2232.
- Dillman, D., 2000. *Mail and Internet Surveys: The Tailored Design Method*. John Wiley & Sons, New York.
- Epstein, D., Avrahami, D., Biehl, J., 2016. Taking 5: work-breaks, productivity, and opportunities for personal informatics for knowledge workers. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, pp. 673–684.
- Feng, J., Lazar, J., Kumin, L., Ozok, A., 2008. Computer usage and computer-related behavior of young individuals with Down Syndrome. In: *Proceedings of the ACM Conference on Assistive Technology (ASSETS)*, pp. 35–42.

- Fogg, B., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., Treinen, M., 2001. What makes web sites credible? A report on a large quantitative study. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 61–68.
- Frankel, M., 1983. Sampling theory. In: Rossi, P.H., Wright, J.D., Anderson, A.B. (Eds.), *Handbook of Survey Research*. Academic Press, New York, pp. 21–67.
- Guy, I., Ronen, I., Zwerdling, N., Zuyev-Grabovitch, I., Jacovi, M., 2016. What is your organization 'like'? a study of liking activity in the enterprise. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, pp. 3025–3037.
- Hanks, K., Odom, W., Roedl, D., Blevis, E., 2008. Sustainable millennials: attitudes towards sustainability and the material effects of interactive technologies. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 333–342.
- Harbach, M., De Luca, A., Malkin, N., Egelman, S., 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)*, pp. 4823–4827.
- Kairam, S., Kaye, J., Guerra-Gomez, J., Shamma, D., 2016. Snap decisions?: how users, content, and aesthetics interact to shape photo sharing behaviors. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, pp. 113–124.
- Kirijian, A., Myers, M., Charland, S., 2007. Web fun central: online learning tools for individuals with Down syndrome. In: Lazar, J. (Ed.), *Universal Usability: Designing Computer Interfaces for Diverse User Populations*. John Wiley & Sons, Chichester, pp. 195–230.
- Kosmalla, F., Wiehr, F., Daiber, F., Krüger, A., Löchtefeld, M., 2016. ClimbAware: investigating perception and acceptance of wearables in rock climbing. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, pp. 1097–1108.
- Lazar, J., 2006. *Web Usability: A User-Centered Design Approach*. Addison-Wesley, Boston, MA.
- Lazar, J., Preece, J., 2001. Using electronic surveys to evaluate networked resources: from idea to implementation. In: McClure, C., Bertot, J. (Eds.), *Evaluating Networked Information Services: Techniques, Policy, and Issues*. Information Today, Medford, NJ, pp. 137–154.
- Lazar, J., Tsao, R., Preece, J., 1999. One foot in cyberspace and the other on the ground: a case study of analysis and design issues in a hybrid virtual and physical community. *WebNet Journal. Internet Technologies, Applications & Issues* 1 (3), 49–57.
- Lewis, J.R., 1995. IBM Computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 7 (1), 57–78.
- Lin, H.X., Choong, Y.-Y., Salvendy, G., 1997. A proposed index of usability: a method for comparing the relative usability of different software systems. *Behaviour & Information Technology* 16 (4/5), 267–278.
- McKenna, K., Bargh, J., 2000. Plan 9 from cyberspace: the implications of the Internet for personality and social psychology. *Personality and Social Psychology Review* 4 (1), 57–75.
- Müller, H., Sedley, A., Ferrall-Nunge, E., 2014. Survey research in HCI. In: Olson, J., Kellogg, W. (Eds.), *Ways of Knowing in HCI*. Springer, New York, pp. 229–266.
- Moser, C., Schoenebeck, S., Reinecke, K., 2016. Technology at the table: attitudes about mobile phone use at mealtimes. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, pp. 1881–1892.
- Normand, L.M., Paternò, F., Winckler, M., 2014. Public policies and multilingualism in HCI. *Interactions* 21 (3), 70–73.

- Nov, O., Naaman, M., Ye, C., 2008. What drives content tagging: the case of photos on Flickr. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1097–1100.
- Ozok, A., 2007. Survey design and implementation in HCI. In: Sears, A., Jacko, J. (Eds.), *The Human Computer Interaction Handbook*. second ed. Lawrence Erlbaum Associates, New York, pp. 1151–1169.
- Ozok, A., Salvendy, G., 2001. How consistent is your web design? *Behaviour & Information Technology* 20 (6), 433–447.
- Pitkow, J., Kehoe, C., 1996. Emerging trends in the WWW population. *Communications of the ACM* 39 (6), 106–110.
- Schmidt, W., 1997. World wide web survey research: benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers* 29 (2), 274–279.
- Schonlau, M., Asch, B., Du, C., 2003. Web surveys as part of a mixed-mode strategy for populations that cannot be contacted by e-mail. *Social Science Computer Review* 21 (2), 218–222.
- Shklovski, I., Kraut, R., Cummings, J., 2008. Keeping in touch by technology: maintaining friendships after a residential move. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 807–816.
- Slaughter, L., Harper, B., Norman, K., 1994. Assessing the equivalence of paper and on-line versions of the QUIS 5.5. In: *Proceedings of the 2nd Annual Mid-Atlantic Human Factors Conference*, pp. 87–91.
- Spears, R., Lea, M., 1994. Panacea or panopticon? The hidden power in computer-mediated communication. *Communication Research* 21 (4), 427–459.
- Sue, V., Ritter, L., 2007. *Conducting Online Surveys*. Sage Publications, Los Angeles.
- Sussman, S., Sproull, L., 1999. Straight talk: delivering bad news through electronic communication. *Information Systems Research* 10 (2), 150–166.