# ·58·

# SURVEY DESIGN AND IMPLEMENTATION IN HCI

A. Ant Ozok
UMBC

## INTRODUCTION

Survey and questionnaire design has been a primary source of data collection within the Human-Computer Interaction (HCI) context since the early days of the science (Baecker, Grudin, Buxton, & Greenberg, 1995). Within the HCI context, surveys are defined as compilations of questions that are implemented either via a computer or paper-and-pencil-based environment, that either have quantitative or qualitative scales, or are open-ended, and that target at extracting a variety of information from a representative sample of the target population (which is in most cases current or prospective users of an HCI system being evaluated).

Survey use is popular in HCI research as it allows researchers to collect, in a relatively easy manner, information based on users' satisfaction, opinions, ideas, and evaluations regarding a system. Design and implementation of surveys are not as costly as conducting experiments in closed environments with special equipment; advances in computer-based survey products and web-based survey services allow direct recording and easy manipulation of survey data by eliminating the need of translation from paper-based to electronic environment; and, with each survey taking minutes to complete in most cases, given a large sample of potential participants can be reached, surveys are a good resource for collecting large amounts of data in a relatively short amount of time and with minimal resources, especially when compared to controlled objective experimental measures that involve in most cases lengthy tasks and recording sessions. On the other hand, surveys are constantly challenged in terms of their validity and reliability mostly due to their high reliance on participant opinions and the impossibility to measure with full reliability that the questions are answered by participants objectively. Quantitative survey research is also sometimes criticized due to difficulties related to survey scaling, as scales rely on an assumption that participants have the same or similar perceptions of scale responses that are subjective in structure (responses such as "I strongly agree" or "I often do it").

This chapter discusses the different aspects of survey design and implementation in HCI in a structured and comprehensive manner. After a discussion of the purpose and a brief history of surveys, the different types of surveys (content and structure-wise), application domains, design, and evaluation techniques are discussed with illustrative examples. The chapter is concluded with emerging and future trends in the HCI survey design and implementation areas.

### Purpose of Survey Use and Survey Applications in HCI

Usability evaluation has been a primary component of HCI since its inception in 1960s. User performance and satisfaction have long been tapped as the major components of usability testing and evaluation (Shneiderman, 1992; Nielsen, 1989). While user-performance measurement relies on objective methods such as software-based time and error measurement, user satisfaction requires more sophisticated tools to be measured objectively. User satisfaction is defined as the level to which a system meets its user's standards and requirements (Hackman & Oldham, 1980).

Directly relating to user satisfaction, user preferences, opinions, and evaluations concerning HCI systems are also of strong interest to usability testing and evaluation processes. Since it is not possible to measure all of these usability components through unequivocal measurement methods, explicit tools have been developed to elicit information regarding user satisfaction, preferences, opinions, and evaluations both qualitatively and quantitatively through user surveys. Surveys serve this specific purpose well by posing targeted questions to users for the purposes of HCI design and evaluation.

While surveys can be designed to collect a variety of types of information concerning the target population, relevant to HCI research and literature (Card, 1996), they are mostly targeted at collecting information in the following three categories:

a. **User Evaluation.** The category aims at collecting information regarding how much a system, product, or environment meets user goals, expectations, and standards. In this category, users are asked a number of questions regarding whether their overall impression regarding the object being evaluated is high, what exactly constitutes this impression, what and where exactly the problems are, and so on. Relating to user satisfaction, this category is also about determining user opinions specific to products or systems, where questions can also include users' opinions concerning whether tasks can be completed effectively and efficiently, whether the system is fast, and so forth.
b. **User Opinion.** The category can, but does not have to, be specific to products, systems, or environments. These types of surveys are aimed at determining what users think about the requirements from a system, product, or environment to fulfill its function satisfactorily. Examples can include surveys that aim at needs assessments for the next generation of cell phones (e.g., what new functionalities can be useful in newer cell phones besides those that already exist according to cell phone customers). Simply put, while the former category consists of surveys regarding the performance of existing systems, environments, and products, the current category is concerned with what users think about what might be useful concerning these systems in more general terms.
c. **Others.** The third category includes the remaining possible survey types aimed at collecting a number of different information types within the HCI context. One such category consists of surveys that are strictly concentrated on population demographics. These types of surveys do not contain questions relying on participants' evaluation of specific products or their opinions, but rather solely on qualifications they own, such as age, sex, education level, skill level, and so forth, or things they do, such as how frequently they go on the Internet or use a cell phone. These types of survey questions are less based on opinion-heavy responses than the previous two categories.

## Brief History of Survey Design in HCI

Surveys started being used as a computer science and, to a limited extent, an HCI research tool in early 1970s, borrowing techniques from anthropology and experimental and social psychology (Myers, Hollan, & Cruz, 1996). With contribution from developments in the overall survey administration and language issues (Converse & Presser, 1986; Belson, 1981; Jenkins & Dillman, 1997), researchers discovered early on that information regarding user attitudes, preferences, and evaluations within the context of computer technology (software and hardware) development can be collected fairly easily with paper-and-pencil surveys. Hence, in the 1970s and 1980s, user surveys were a part of computer research with a social psychology emphasis, but not directly relating to usability testing and usability-design issues. In the mid-1980s graphical user interfaces became an important part of end user computing and usability research took off. The surveys gained a more significant role in HCI research around the same time, and with the advent of graphical user interfaces, surveys in HCI and specifically usability research gradually gained importance. With the graphical user interfaces as we know today gaining high popularity with Windows 95, usability research accelerated (Myers, 1996), and besides building usability laboratories, companies and research institutions started developing and implementing surveys to determine user trends and preferences in HCI. Advanced electronic and paper-and-pencil survey-design methods have been developed in the last decade (Dillman, 2000), and user surveys have become an essential part of data collection in HCI research.

## Paper-and-Pencil and Electronic Surveys

Survey implementation largely relies on practical factors. Besides the challenge of finding a sample size that is both large enough and representative of the population the study is targeted at, implementation challenges include presenting the participant sample with a survey that is quick and easy to complete, has a fun factor and a learning component for the participants, and does not require tedious data extraction and manipulation once implementation is completed. Additionally, surveys in every topic should be unambiguous, unbiased, correctly coded, piloted, and ethical (Stone, 1993). Today's surveys are almost universally implemented in two forms: (a) paper-and pencil surveys, which require participants to mark or write their responses on response sheets that also contain questions, either on the same sheet or separately; and (b) electronic surveys, which require the users to use the screen, keyboard, and/or mouse of a computer to mark or type their responses on the screen.

Paper-and-pencil-based surveys require the survey material to be physically distributed, filled out, and returned. This process can occasionally be cumbersome and tedious. Moreover, these types of surveys also require manual entry of quantitative, and in most cases qualitative data, for analysis. One solution to the problem of translation of paper-based data into electronic format is using Scantron sheets which are sometimes referred to as "bubble-sheets." In this system, designated areas (bubbles) on printed sheets are filled out by participants with a pencil, and these sheets are then fed into a computer and scanned for correct answers. This process, however, is quite costly due to the scanning equipment necessary for the process. Figure 58.1 shows two sample paper-and-pencil survey sheets, one with response spaces below each question and one with response sheets that are separate from question sheets.

Although about 62% of all American households own one or more computers (U.S. Census Bureau, 2005), computers are still not accessible to the entire population. Therefore, paper-and-pencil surveys are still widely popular. Paper-and-pencil surveys allow swift collection of large data quantities if they are administered to groups of participants simultaneously, such as a group of students during a class period who are asked to fill out and return the surveys immediately. One other common way of paper-and-pencil-based survey implementation is mailing in surveys to participants and asking them to return them via mail, in most cases in postage-prepaid envelopes. However, recent studies indicated that return rates of mailed-in surveys by the participants is highly variable depending on the type of survey (Dillman, 1991, 2000). Taking into consideration the percentage rate of computer ownership among American households, mailed-in surveys can be concluded as a less-preferred means of data collection specifically within the context of the HCI research, and mail surveys are therefore not popularly used for HCI research purposes.

Computer-based surveys (sometimes referred to as PC-based surveys) have become popular with the advent of home computers in the 1990s, even before the high adoption of the World Wide Web. In computer-based surveys, participants are presented the survey on a specific, standalone (non-networked) computer. It should be noted that these types of surveys are not web-based, but rely on the software installed on the computer on which the survey is implemented.

Participants use the mouse to click on their responses of choice for multiple-choice questions. Mouse actions are generally implemented on dropdown combo boxes (where a dropdown menu opens up with the options when the user clicks on the button located to the right side of it) or radio buttons (a round area to the left of the option is clickable), or check boxes where multiple choices can be clicked on at one time (a square-shaped box to the left of the option is clickable) (Ozok & Salvendy, 2000). For text entries, participants can type text on specified text boxes. While computer-based surveys can be convenient because of having the initial data in electronic format and eliminating the necessary transformation to electronic format in paper-and-pencil-based surveys, they require the participants to be stationary on a specific computer. For large-size implementations, computer-based surveys can be extremely slow in collecting the data, mainly due to limited computer equipment and scheduling difficulties. It can be concluded that while computer-based surveys can be advantageous in the data-analysis stage and are still popular in kiosks stationed in public places such as shopping malls, they are not suitable for large-sample size and lengthy surveys, and hence are not the best solution in survey-based data collection in HCI context.

With the advent of the Internet, web-based (online) surveys have become highly popular (Dillman, 1999) and allowed researchers to conduct surveys more effectively and efficiently than

1. My cell phone screen is convenient to use.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

2. The keypad on my cell phone is a convenient means of data input.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

3. My cell phone interface is easy to use.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

**Question Sheet:**

1. My cell phone screen is convenient to use.
2. The keypad on my cell phone is a convenient means of data input.
3. My cell phone interface is easy to use.

**Response Sheet:**

| | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|---|
| 1. | Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |
| | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| 2. | Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |
| | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| 3 | Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

FIGURE 58.1. Two paper-and-pencil presentations of the same survey, one with the response spaces below each question and one with separate question and response sheets.

more traditional means (Zhang, 2000). Server-based software allows survey participants to access a web page and fill out the survey, then submit their results mostly to a central server by clicking an on-screen button. web-based survey interfaces in structure look very similar to computer-based surveys with same interface elements of dropdown combo boxes, radio buttons, check boxes, and text boxes serving the functions of various data-entry types by the participants. The data are collected on a central web server in these types of surveys, which can be easily obtained and manipulated by the survey administrators. Additional data storage and analysis programs residing on these web servers can compile the data in a variety of formats such as Microsoft Excel, and also implement some automatic data analyses such as calculation of means and standard deviations (descriptive statistics).

One additional electronic survey type consists of the administration of e-mail surveys in which participants are e-mailed a survey and are asked to fill it out and e-mail it back. However, with the dramatic increase in the amount of spam e-mail users receive in recent years, it can be concluded that these kinds of e-mails are likely to be perceived as spam and are likely to be discarded. Therefore, e-mail surveys are not articulately covered in this chapter.

There are two types of methods used in web-based survey administration today. One method is to use a web survey administration service provider (such as SurveyMonkey.com) by paying it a monthly or yearly fee. In most cases, various packages are available ranging from a small number of surveys to unlimited administration of surveys. The web service providers also have a variety of options for the retrieval of survey data by the administrators, for example in Access or Excel formats, with some, in most cases basic, statistical analyses (such as descriptive statistics and correlations) already performed on the data. Additionally, the services also give flexibility in customization of survey interfaces such as giving the freedom to the administra-

tors (and in some rare cases to the participants) to choose their text and background colors, font sizes and types, how many questions to have per web page, and so on. Today, there are more than a hundred online survey-service providers with monthly fees varying from $3 for small-scale academic surveys to $500 for large-scale, mostly marketing-related surveys.

Another means of administering online surveys is to use one's own hardware and software. As a central server is necessary for collection and storage of the data, this hardware can either be purchased or rented, or an existing server can be used. The amount of storage space necessary largely depends on the length and sample size of the survey, but since most surveys contain textual data, it is almost unimaginable to need more than five gigabytes of storage space for a large-size survey for purposes of HCI research. A variety of open source (such as php ESP [Easy Survey Package]) and licensed software vendors (such as Inquisite) are available for survey implementation and administration on administrator-owned servers. Like survey-service providers, these software packages also allow a variety of customization flexibilities concerning the survey interfaces. Figure 58.2 presents a sample interface from an online survey.

For both paper-and-pencil-based and electronic surveys, human cognitive limitations should be taken into consideration and basic human factors guidelines should apply. Paper-and-pencil surveys should be administered with pencil to allow participants to go back and change their responses if they want to. White paper with black, 12–14-point-sized Times New Roman font text should be used, as those font sizes are the most common and easily readable text sizes in printed documents. Labels should be presented with bold, 16–18-point-size text, and while pictures are seldom presented to participants, if they are presented, they should have enough resolution to appear clearly on paper. In short, paper-and-pencil-based survey interfaces should be inspected to make sure they comply with the structural human factors guidelines for paper-based interfaces.

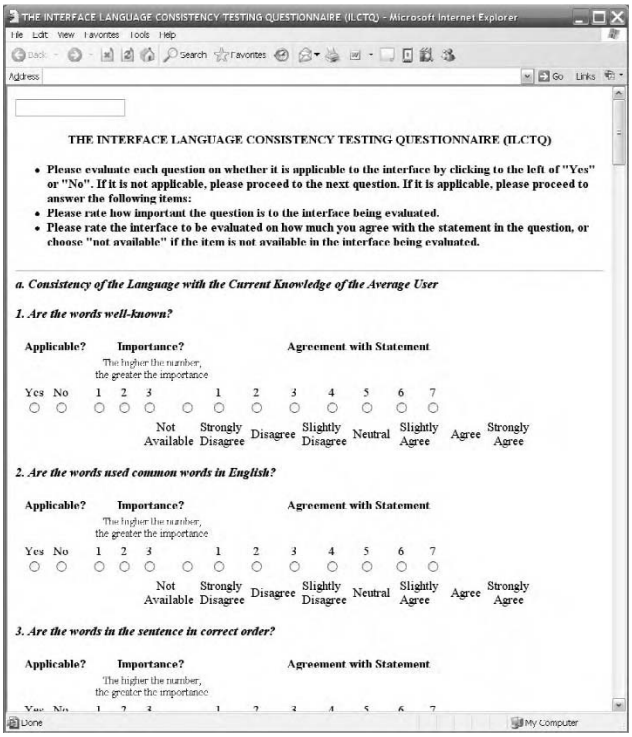


FIGURE 58.2. Sample interface from an online survey.

Similarly, electronically administered surveys (web- or PC-based) also should follow basic human-factors guidelines for computer interfaces. For web-based surveys, it should be noted that participants will access the surveys from a variety of different types of machines, and basic computer and web design guidelines indicate 800×600 color screen resolution is the most common screen type (Nielsen, 1993; Badre, 2002) which should be taken into consideration when survey interfaces are designed, making sure that survey objects, mainly text fonts and some images if there are any, are easily visible on screens with this resolution. The basic rule of black text on white background in web usability should also be applied in computer- and web-based interfaces, and screen objects such as dropdown combo boxes, radio buttons, buttons, and check boxes should be the same size (mainly height) as text lines to ensure consistency and easy visibility and to allow users to click on them easily. Text boxes for the users to type in should allow text size consistent with the text on the question parts of the interface. Overall, again, it can be said that in web-based surveys, basic web-design guidelines can easily be adopted.

Additionally, one important item in survey design is to prevent survey participants from getting worried about the excessive length of the surveys, as too-long surveys may result in significantly decreased response rates (Lund & Gram, 1998; Krosnick, 1999). Taking also into consideration the fact that computer users don't enjoy scrolling down the screen in general (Shneiderman, 1992), no more than 20 questions should be presented on one survey screen. For surveys containing more than 20 questions, subsequent questions should be presented on subsequent screens (again, each screen not exceeding 20 questions), which participants should move to by clicking on a screen link or button that should have a statement such as "Click here to continue."

Both paper-and-pencil-based and electronic surveys will continue to be used in HCI research. While it is unlikely that the former will go extinct any time soon, recent studies (such as Dillman, 2000) showed that web-based surveys have very significant advantages in data collection and analysis stages of survey-based research. The fact that most HCI-related survey research uses computer-savvy sample participant groups is also a factor that helps the popularity of web-based surveys within HCI. It is therefore expected that web surveys will eventually constitute to a large majority of survey-based research in HCI. While no previous study explicitly investigated the exact share percentage of paper-and-pencil-based and electronic surveys in HCI research, it is estimated that more than 60% of all surveys for HCI research are implemented on the web environment.

While the discussion in this chapter mostly concentrated on structural issues of paper-and-pencil-based and electronic survey design, a much larger research topic concentrates on how the actual survey content should be designed, created,

and implemented. The next chapter discusses content creation in surveys within the context of HCI research.

## SURVEY DESIGN IN HCI

A major part of survey-design research in HCI is concerned with the formulation of survey questions and scales. In comprehensive user surveys, questions may be presented as one large, continuous list. Related questions may also be presented in categories and sections, such as the demographic questions in the first section, questions evaluating the structural elements of the interface in the second section, and so on. As discussed earlier, surveys in HCI research are mostly concentrated on collecting two categories of information: participants' opinions and evaluations. These two survey types are discussed in the next section. This discussion is followed by an explanation of survey application domains and the survey design methodology.

### Survey Types

#### *Opinion Surveys*

Opinion surveys aim at determining what participants think about issues pertaining to certain concepts. Additionally, opinion surveys can also measure participants' wishes, habits, and customs (Baker, 1998). The way opinion surveys differ from evaluation surveys is that opinion surveys are not centered on a specific product or environment, but are more general in nature. For example, an opinion survey can target measuring cell phone users' habits and opinions regarding the use of the cell phone interface (whether they are happy with the screen size, whether they can easily enter data using the keypad, etc.). While questions are focused on a product in this example (e.g., a cell phone) the population of cell phone users is highly diverse, as are cell phone devices that are used by the population. Hence, such an opinion survey will measure general trends in usability concerning cell-phone interfaces rather than evaluating a specific cell phone interface.

Opinion surveys are administered with more freedom than evaluation surveys, as the participants do not have to have the evaluated issue, product, or environment fresh in their memory. Hence, they can fill out opinion surveys at any time and in any environment that is convenient for them. Opinion surveys include questions that do not require any recalling process, meaning they contain questions which participants can answer without having to recall a specific feature or part of the environment, product, or issue. As opinion surveys ask general questions about participants' current standing regarding opinions, wishes, customs, and habits, the answers may sound subjective and vary greatly among participants. Sample questions for opinion surveys include statements such as "Does the use of computers in daily tasks contribute to your overall technology knowledge?" or "Are you in general satisfied with the amount of product information available on an e-commerce page you frequently visit?" or "How would you rate the customer services on the sites where you frequently shop?"

Opinion surveys can cover a broader variety of issues than evaluation surveys which are more focused. They can include both qualitative and quantitative scales for their responses. Although no previous literature came up with a strict classification of opinion surveys, the following classification can help in structuring of opinion surveys and what kind of questions should be asked for what types of opinion-related survey categories in HCI research (Sample questions for each classification of opinions are presented on Table 58.1):

a. **Opinions on a medium.** Within the context of HCI research, these types of surveys concentrate in most cases on interface design, usability, user satisfaction, and user preferences issues concerning a medium the participants use in their daily lives. Most popular examples of this type of media include daily-used devices such as computers, Personal Digital Assistants (PDAs), and cell phones. These types of surveys concerning users' opinions on a certain medium aim at determining general trends in user opinions concerning whether the design of the medium is satisfactory and meets user needs and requirements, whether there are problems concerning the medium, and what can be possible solutions. Opinion surveys concerning a medium are also useful when they are used by usability specialists and engineers to develop new versions of products or to come up with new products, as the survey results can pinpoint the needs of the target population to be met regarding the medium. Sample questions for these types of surveys can include questions like "Are you in general satisfied with the web browsing capabilities of your cell phone?" or "What additional capabilities would you like to see on your PDA user interface?"
b. **Opinions on an event.** Within the context of HCI research, user opinions on an HCI-related event can include what they think about certain HCI-related happenings. Examples can include opinions concerning certain HCI-related activities, with questions like "Do you find the move from a text-based interface to a graphical user interface helpful?" These types of surveys are rarer in nature and aim at collecting basic trends concerning user opinions in current HCI-related activities.
c. **Opinions on a procedure.** These kinds of surveys aim at determining the user opinions on procedures to complete HCI-related tasks. They are similar to medium-related opinion surveys, but rather than questions about the medium itself, these surveys have the goal of determining user opinions on how it is used. In web and e-commerce design, these kinds of opinion surveys are helpful in determining whether the procedures to complete general tasks (for example, web navigation) meet user requirements and needs. A sample question in an opinion survey concerning a procedure could be "Are you satisfied with how long it generally takes to purchase a product on an e-commerce site?" Surveys to explore opinions on a procedure are less common as HCI researchers usually resort to evaluation surveys to test procedures for use of computer interfaces and other media (explained in the next section).

TABLE 58.1. Opinion and Evaluation Survey Sample Questions

**Opinions on a Medium:**

1. Personal Digital Assistants (PDAs) are convenient.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

2. I would prefer using a laptop to a desktop computer for my daily tasks.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

3. Tablet PCs are capable to meet my daily computing needs.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

4. Setting up a wireless Internet connection is easy on a computer.

| ○ | ○ | ○ | ○ | O | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

**Opinions on an Event:**

1. Free wireless area networks should be universal.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

2. Advancements in Liquid Crystal Display (LCD) technology made them more affordable.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

3. Free services on the Internet meet my expectations.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

4. The Windows computer interface is significantly easier to use now than it was five years ago.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

**Opinions on a Procedure:**

1. The sign-up process for e-commerce sites is too time-consuming.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

2. Password authentication procedures are reliable.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

3. The steps involved in personalizing a cell phone interface settings are long.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

4. Payment transaction procedures on the Internet are mostly secure.

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

**Evaluation (of a Computer Interface):**

1. How difficult did you find completing the tasks on the interfaces?

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Extremely Difficult | Difficult | Somehow Difficult | Neither Easy Nor Difficult | Somehow Easy | Easy | Extremely Easy |

2. How intuitive were the interfaces you used?

| ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|
| Not Intuitive at All | Not Intuitive | Neutral | Intuitive | Very Intuitive |

3. While completing the tasks, I made _____.

| ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|
| No Errors | Few Errors | Neither Many nor Few Errors | Quite a Few Errors | Many Errors |

4. Were the interfaces aesthetically pleasing?

| ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|
| Not at All | Very Little | Neutral | Quite a Bit | A Lot |

### *Evaluation Surveys*

More specific than opinion surveys, evaluation surveys (or questionnaires) are generally administered after a certain procedure is implemented on the participant group. While opinion surveys can be administered at any time to the participants, evaluation surveys are administered right after the participants have completed a certain procedure. In evaluation surveys, participants are asked about tasks they have just recently completed. Therefore, evaluation surveys are in most cases preceded by certain usability-related experimental procedures. Most common in HCI, evaluation surveys are administered after participants have completed a number of tasks in controlled computer environments. They are also implemented right after the procedure to ensure that memories regarding the procedure are still fresh in the participants' minds, as evaluation

surveys require a high amount of recall of procedures and interfaces in the tasks that were just previously completed.

HCI-related evaluation surveys have the main goal of evaluating usability, user satisfaction and user preference issues concerning user interfaces or environments (Ozok & Salvendy, 2001). After certain tasks are completed in these controlled environments, evaluation surveys aim at determining the exact nature and location of problems and points open to improvement in the human-computer environments. Therefore, evaluation surveys are in most cases relatively detailed in nature. In a sample procedure for example, participants can be presented a number of web pages and asked to complete some common user tasks on those pages, tasks that can include text entry, direct manipulation, and form filling. An evaluation survey that would follow could include questions such as "Was the site navigation difficult?", "Was the text size on the site readable?", "Did you enjoy the overall experience on this site?", "Would you visit this site again?", and so on.

As their name indicates, evaluation surveys aim at evaluating interfaces, environments, and procedures from the user's perspective. For that purpose, they are tools to determine how participants evaluated those interfaces, environments, and procedures. In that sense, evaluation surveys are explicit and not that much different from objective measurement methods such as performance measurement in HCI tasks. Similar to opinion surveys, evaluation surveys use both qualitative and quantitative scales.

Evaluation surveys are not only helpful in evaluation of product interfaces, environments, and procedures. They can also be used in evaluating certain conceptual features. For example, a researcher may be investigating whether the introduction of a certain interface concept results in higher user satisfaction regarding the interface. If this particular feature is, say, that of interface visibility, the survey administrator can first present the participants with high-visibility and low-visibility computer interfaces, and the subsequent evaluation survey can contain questions such as: "Did higher visibility improve your overall satisfaction with the computer screens?"

Evaluation surveys are useful in a variety of HCI-related commercial and academic research activities. In academic research, different computer or computer-related interfaces can be evaluated through surveys to determine whether they result in better user preferences or higher user satisfaction. In commercial research, newly developed products (for example a new computer peripheral), environments (for example a new computer interface), or procedures (for example the steps it takes to complete a transaction with a new e-commerce site design) can be empirically evaluated by having participants complete tasks with those products and procedures or in those environments, then filling out evaluation surveys consisting of detailed questions regarding their satisfaction and preferences regarding the said product, environment, or procedure. While quantitative evaluation results can give statistical backing to user evaluations, helping boost their conclusiveness, qualitative evaluation results can give the researchers who administered the surveys new ideas to improve usability/user-preferences-related design components. Therefore, evaluation surveys are the most commonly used survey types in HCI research and constitute to one of the most common and effective user evaluation methods in HCI in general. Examples of opinion survey questions are also presented in Table 58.1.

### *Other Survey Types*

Besides the two main survey types mentioned earlier, one widely used survey type is the demographic survey. Although demographic surveys are almost universal, it would be incorrect to categorize them at the same level as opinion and evaluation surveys, as most user surveys have a section concerning user demographics. Hence, in most cases demographic survey are essential parts of opinion and evaluation surveys rather than stand-alone surveys. In some cases HCI researchers administer surveys consisting of demographic questions only—for example, to determine the demographics of a user group using a specific application. However, more commonly, HCI surveys consist of opinion and/or evaluation questions in addition to the demographic questions.

Demographic questions play an important part in HCI-related survey design, as most variables of interest in HCI research are also dependent on the factors that are specific to the target population. Consequently, research findings can only be generalized in most cases to the target population from which a representative survey is sampled.

Demographic surveys (or survey sections) in most cases consist of a standard set of categories: age, sex, education, and occupation of the participant. Age can also be asked in the form of "birth year," and for the question regarding sex, options in the form of "Male/Female" can be presented to participants to mark on the computer or with the pen/pencil. Education level can be left to type in or write, or options can be presented. The question is usually formulated as: "What is the highest degree for which you won a diploma?" or "What is the highest degree you earned?" Typical options for this question are "Elementary School," "Middle School," "High School," "College or University," "Graduate Degree," and "Post-Doctoral Graduate Degree." The occupation question is about what kind of job the participant has. For this question, usually the participants are asked to type or write in the designated area, due to the high variety of possible occupations participants may have, although presenting options for this question is also possible if some general occupation categories are all that is needed, for example options such as "Private sector," "Academia," "Student," and so on.

In addition to this basic set of survey questions, demographic surveys can also include general questions regarding daily habits or current standing issues concerning the participants. Most commonly, demographic surveys in the HCI area contain questions regarding computer use habits of participants, such as "How many times a day do you check your e-mail?" or "How many times in the last year did you shop from a web-based e-commerce company?" These types of questions are usually customized according to the type of information needed for the specific research being conducted and can therefore greatly vary in nature, but in principle they aim at collecting information on computing-related habits in most cases. Table 58.2 presents a set of sample demographic questions as part of a survey regarding cell phone use.

TABLE 58.2. Sample Demographic Questions from a Survey on Cell Phone Use

| |
|---|
| Your Age: |
| Your Gender: |
| Your Occupation: |
| How Many times a week do you go on the web?: |
| _____ Less than once a week |
| _____ Between once and three times |
| _____ Between three times a week and every day |
| _____ Every day |
| In the past year, how many times did you shop online (Please put a number)?: ____ |
| Do you own a cell phone, a Personal Digital Assistant, or a Combination Device? |
| _____ Yes |
| _____ No |
| In the past one year, how many times did you shop online using a cell phone, a Personal Digital Assistant or a Combination Device (Please put a number)?: _____ |

It should be noted that demographic questions are of a more personal nature than opinion and evaluation questions. Some participants may feel that their privacy rights are being violated by being asked to provide their age and education level. For this reason, providing anonymity in surveys and informing the participants about their provided information being not personally identifiable—in other words, providing anonymity—is greatly crucial in HCI research. Knowing that their data cannot personally identify them usually takes care of privacy worries and is known to improve participant enthusiasm. Challenges concerning privacy in survey implementation are described later in this chapter.

There are no other significant survey types widely used in HCI research. Some niche survey types may still exist but are few and far between. Therefore, surveys aimed at collecting information relevant to HCI research usually belong to one of the categories explained in this section. In the next section, application domains of surveys relating to HCI research are discussed.

## Survey Application Domains

Survey applications are highly popular in a broad range of application domains, in areas ranging from social sciences to marketing, education to production, customer to worker satisfaction, and many more. Today, results obtained from surveys which in most cases ask comprehensive questions are deemed reliable and valid in both scientific and industrial projects. The most common application domains of survey research include

a. **Sales and Marketing:** Companies that offer products and services for both consumers and industries use customer/client surveys for both needs-assessment and evaluation purposes. A large number of companies are also solely dedicated to implement customer surveys for companies, analyze the data, and deduct conclusions for sales and marketing purposes. Customer satisfaction, product evaluation, customer relationship management, and customer demographics are only a few of the topics surveyed by sales and marketing forces.
b. **Medicine:** Medical research is not limited to trials relying on objective measurements. Surveys can be helpful in collecting patient data for development of medicine or treatments.
c. **Education:** Educational surveys can help determine population preferences in education as well as education levels and education-related difficulties among population segments.
d. **Information Technology Research:** In the field of information technology, surveys are widely used in connection to software and hardware design and evaluation, covering a broad variety of areas including software engineering, systems analysis and design, and of course human-computer interaction, which this chapter covers.

The earlier-mentioned relevant list covering the application domains of surveys is far from complete, but a sample of application domains are presented in the list. The list will no doubt continue growing with the advent of new technologies and sciences. The HCI area is seen as a major application domain of surveys, and is expected to continue to be so.

## Survey Design Methodology

Survey design is a methodological activity that requires a systematic design procedure. Survey design techniques and procedures are discussed in this section, including content design and scale design, followed by survey design and redesign issues, a survey design example, and a discussion of challenges in survey design.

### *Survey Design Techniques*

Survey design mainly consists of two components: the design of the survey content and the survey scale. They are both discussed in this section.

***Content design.*** In the heart of the survey research lays the issue of producing the actual questions to ask the participants. Designing the survey content is actually producing these questions along with their scales. Deciding on which questions to ask to the participants in the survey largely depends on three resources: literature, expert opinions, and individual experiences.

A large number of survey questions are based on previous research in the focus area. Relying to some extent on previous literature allows the researchers to achieve high validity of their survey structure, as previously validated research allows current survey design to have strong backing in terms of its content and the targeting of the questions concerning the particular

research topic. Therefore, it is best to have backing from previous studies at least for the majority of the questions while designing the survey.

While a designed survey's content may consist largely of questions that are based on the relevant literature in the area, there will be most likely some issues that are intended to be included in the survey but are not covered in the previous literature. Therefore, HCI researchers sometimes rely on experts in the area to cover additional points to be included in the survey. A preliminary survey may be sent in this context to the area specialists to determine the most significant items to be covered in the investigated area. For example, if a survey research is trying to determine the most significant interface design items in e-commerce that affect buying behavior, a preliminary survey may be sent to experts in the area (for example, e-commerce company managers and professors in business schools specializing in e-commerce) to determine the general classifications of interface issues relating to the buying decision. In addition to the literature, these responses from experts can be used as a major resource of question generation for the resulting survey.

Researchers can also rely on their own heuristics and expertise in producing questions. To prevent being accused of "making up the questions," the researchers would need to explain logically why the questions were included in the survey. In these cases, researchers can include questions based on what they think is a significant part of the research item being investigated, or based on the impression that, although the literature did not explicitly point out the issues in these types of questions, there was an implicit indication in previous research towards this particular direction.

Design of survey content is not a difficult task once the researcher has a reasonable background in the area of interest. One common mistake done in design of survey content is the researchers missing important questions during the design and ultimately not addressing those questions. Therefore, cautious, repeated reviews and revisions are necessary before the final implementation of the survey.

***Scales and open-ended questions.*** Just as important as the content of the questions, the scales for the survey questions in HCI are essential for the accuracy and validity of survey results. Scales are created to attribute numerical values to participant responses in the survey, thereby allowing statistical analyses and giving statistical backing to conclusions obtained from the research. To respond to a scaled question, the participant marks one of the several options, the option which best represents his or her opinion regarding the item in the question. If the question has a large variety of possible answers, or if it requires a lengthy answer, then an open-ended response style may be preferred rather than presenting a scale to the participant. For open-ended responses, participants are mostly given the freedom to write or type as much as they would like.

Both scaled and open-ended questions are suitable for different question types and types of information being obtained from the participants in the survey. Quantitative studies have to use numerical scales to statistically test their hypotheses and support their findings. Qualitative research, on the other hand, analyzes survey data without the involvement of numbers. Because participants have a much higher degree of freedom when responding to open-ended questions, qualitative responses are not restricted to the defined response universe determined by survey designers (also referred to as "survey authors"). On the other hand, conclusions derived from qualitative responses may be more arguable because they cannot be tested statistically.

One more type of responses in HCI surveys includes participants being given the freedom to mark more than one response choice. While these types of responses are generally not assigned numerical scales, these types of responses are presented in demographic surveys. In these types of questions with possible multiple responses and in open-ended questions, it is useful to present the option of "Other (please specify):" to the participants, as there is always a possibility that the survey designers may not present the option which the participant would like to give as a response. A sample question of this sort could be "Where do you generally access the Internet?" with the possible responses "Home," "Work," "School," "Coffee Shop," "Internet Café," and "Other (please specify)."

It should be noted that one alternative to open-ended survey questions are interviews and focus groups, and these more interactive data-gathering techniques are likely to result in the collection of richer data than open-ended survey questions as they allow real-time interactions between the researchers and participants. Therefore, it is not highly common in HCI research to use surveys with open-ended questions only. In most cases, especially in quantitative survey research, a mix of both open-ended and scaled questions often proves to provide best empirical results. Due to their higher frequency of use, this book chapter is more focused on the design and implementation of scaled surveys rather than surveys with open-ended questions. As part of this direction, scale design is discussed in the next section.

***Scale design.*** While a large variety of scaling techniques are available for surveys in sociology and psychology research, HCI surveys mostly rely on Likert scales (Medsker & Campion, 1997). While contrast scales consisting of yes-or-no questions with 1/0 corresponding scales are also used, five and seven-point Likert scales are highly common (Aiken, Lewis, & Lewis, 1996). In most cases, a scale needs to consist of and odd-number of options. This way, the option in the middle can correspond to "no preference" or neutral" opinion (Dillman, 2000). Each response on the scale attributed a number to allow the researchers to conduct statistical analysis on the collected data. Item scales need to be kept consistent during the data analysis phase, meaning items should be lined up in the same direction, whether they are positive or negative—in most cases positive responses scoring high and negative responses scoring low on the scales. Inverted questions (questions that ask items in the opposite direction, as discussed later in this chapter) should have their scales reversed in order to keep consistency and allow correct data analysis. A large amount of data analysis mistakes in surveys usually happen because of scaling problems. For example, if the researchers forget to invert scales of reverse questions, then correlations and differences between responses will not come out correctly, resulting in lack of validity of research conclusions.

Scales can indicate a number of different issues. Some scales are concerned with user opinions while others are concerned about frequencies. Most common scale types include agree-

ment measurement ranging between "Strongly agree" and "Strongly disagree"; frequency measurement ranging between "Not at all" and "Very often"; quality opinions ranging between "Very good" and "Very poor"; opinions regarding probability ranging between "Very unlikely" and "Very likely"; and so on. It should be noted that survey scales offer a certain amount of freedom to survey designers on how to name the possible response options for their questions, and therefore scales come in many different varieties, from those measuring amounts (a lot, quite a bit, etc.) to frequencies (very often, often, etc.) to yes-or-no scales. Normally, in most cases a "Not Applicable" or "Not Available" option needs to be added to the scale. When this option is marked, this particular question of this particular subject is eliminated from the analysis. Note that this is different from giving a score of zero to that question in the analysis. Table 58.3 present a sample of possible response scale sets along with possible number correspondences to the responses.

### *Survey Evaluation Techniques*

After the initial design of the survey questions, scales, and instructions to the survey participants on how to fill the surveys out, surveys need to be evaluated to determine whether they are measuring what the designers intended them to measure, whether they are reliable, and whether they produce valid results. Pilot testing is one common method to preliminarily address these issues. However, full evaluation of a survey can mostly happen only after substantial data have been collected with the survey as a tool. Therefore, the evaluation of a survey is conducted based on the data collected by it, and the two components of survey evaluation are the measurement of the survey's validity and reliability.

TABLE 58.3. Possible Survey Responses and Their Numerical Equivalences

| | | | |
|---|---|---|---|
| *Strongly Disagree* | 1 | ***Never*** | 1 |
| *Disagree* | 2 | ***Very Seldom*** | 2 |
| *Moderately Disagree* | 3 | ***Seldom*** | 3 |
| *Neutral* | 4 | ***Neither Seldom Nor Often*** | 4 |
| *Moderately Agree* | 5 | ***Somewhat Often*** | 5 |
| *Agree* | 6 | ***Often*** | 6 |
| *Strongly Agree* | 7 | ***Very Often*** | 7 |
| | | | |
| ***Excellent*** | 5 | *None* | 0 |
| ***Good*** | 4 | *Very Few* | 1 |
| ***Fair*** | 3 | *Few* | 2 |
| ***Poor*** | 2 | *A Fair Amount* | 3 |
| ***Very Poor*** | 1 | *Quite a Bit* | 4 |
| ***Not Applicable*** | — | *A Lot* | 5 |
| | | | |
| *Not Convenient at All* | 1 | ***Very Difficult*** | 1 |
| *Highly Inconvenient* | 2 | ***Difficult*** | 2 |
| *Inconvenient* | 3 | ***Not Difficult*** | 3 |
| *Neutral* | 4 | ***Easy*** | 4 |
| *Somewhat Convenient* | 5 | ***Very Easy*** | 5 |
| *Convenient* | 6 | *Yes* | 1 |
| *Highly convenient* | 7 | *No* | 0 |

***Survey validity and validity determination.*** While the reliability of a survey is determined after the survey data has been collected, the validity of the survey has to be determined prior to the implementation. As the name implies, validity of a survey is the degree with which the survey instrument is valid in the results it produces, in other words whether the survey is measuring what it says it is measuring (Litwin, 1995). Generally, within the context of HCI, validity is covered twofold: *construct validity* indicates the degree of how much the survey is backed by previous research in its field, how solid its construct is. In general, as in every research, development of a survey needs to rely on previous research to give the tool literature backing, proving that the survey didn't come out of the imagination of the designer, but rather relies on a number of different research studies conducted by a number of different researchers. To prove the construct validity of their survey, designers need to prove the case that the questions they put into their survey are based on previous literature. Hence, in survey design, it is imperative to ensure that a majority of the questions have been implicated in the previous literature as items relevant to the current topic of interest. Without being able to prove this validity, it is not possible to make a convincing case regarding whether the survey is doing an undisputed contribution to the overall research topic of interest. However, it should be noted that it is almost impossible to provide a survey in which every single item has a full set of articles or books backing it. In most cases, some survey questions may have some indirect mention in the previous literature, and some survey questions may be solely based on the individual experience and/or opinion of the survey designers. This type of question generation is also acceptable, as long as the designers can prove that those questions are also based on solid research. In short, construct validity of a research survey in HCI aims at proving the conclusion unarguably that results obtained from this survey are on target and valid. Validity is therefore crucial to the success of the research conducted, of which the survey is a part.

Predictive validity is, simply put, the ability and power of the survey to predict correct results in repetitive use. A survey with predictive validity indicates that the results obtained from it in the current and future use have the power of predicting accurate results. For example, if a comprehensive survey has been produced measuring the usability of a website, researchers will need to prove as part of their study that once this developed instrument is administered, the results that are produced accurately reflect the usability level and attributes of a website. Additionally, the survey also needs to accurately reflect usability levels and attributes when it is administered on other participant groups for evaluation of other sites. If these capabilities of the survey can be proven by the researchers, then the survey can be said to have predictive validity.

Like construct validity, predictive validity does not have any quantitative measurement method to be used. Hence, a survey's predictive validity again relies on qualitatively proving that the survey results are based on solid research notions, and hence the results are accurate. Survey designers need to explain that the results produced from their surveys have been proven to accurately reflect current situations concerning the target population's specifications, evaluations, and opinions, and will continue to do so in future applications when it is administered

again. For this purpose, the elements of the survey need to be proven as elements that accurately predict results concerning the topic in focus. To make accurate predictions, surveys need to consist of elements that make accurate predictions themselves when their results are analyzed, and to ensure that these elements have predictive power, they need to rely on accurate literature findings and real-life situations. In short, similar to construct validity, predictive validity of a survey can be accomplished by ensuring that the survey relies on solid previous literature and the researchers' findings. Sometimes, to ensure that survey findings and questions have predictive power, they are evaluated by experts in the area prior to the implementation. Pilot testing is addressed later in this chapter.

***Survey reliability and reliability measurement.*** Reliability of a survey is the measure of whether the survey is measuring things consistently, and whether the results obtained from the survey can be relied upon. A survey's reliability affects its validity, as a survey that is not reliable in its measurements cannot produce fully valid results.

While there are a number of quantitative reliability measurement techniques for survey design, especially in psychology and sociology, the two most common reliability measurement techniques used in HCI research are *internal* and *inter-rater* reliability techniques.

The internal reliability technique is concerned with whether the survey questions are understood by the participants the way they are intended to be understood when they were prepared by the survey designers. An internally reliable survey contains questions that are all comprehended the same way by all participants at all times in repeated measures when it is administered. A lack of internal reliability is a common phenomenon, as different participants can understand survey questions differently if they are not asked in a highly clear and unambiguous fashion. Therefore, to improve the internal reliability of surveys, designers need to make sure to use statements that are entirely clear and leave no room for interpretation on what is meant in the questions. An example of a low internal reliability survey question would be "Did you have tremendous difficulty completing the tasks on the web page?" In this question, participants who had little difficulty, no difficulty, and a moderate amount of difficulty may respond to the questions in a very similar way, resulting in confusion regarding whether the tasks were difficult or not. Additionally, if participants had difficulties in some tasks and no difficulties in the others, a question such as this may confuse the participants on what types of tasks (difficult or not difficult) they should base their response on. Obviously, in survey design it is important to be careful not to confuse the participants while they are filling out the surveys. Potential confusions can mostly occur on the participants' parts regarding what is meant by the survey question, and what the survey question is about (Cronbach, 1990). Surveys may have high construct validity, meaning they may have been designed based on solid research, but if they confuse the participants with their questions, they will obviously lack internal reliability, and consequently, predictive power.

The most commonly used measure for internal reliability of surveys is called "Cronbach's Alpha Internal Reliability Coefficient" (Cronbach, 1990). The coefficient relies on checking whether participants respond to the same question the same way when it is asked the second time, in a similar form. These types of questions are called "duplicate questions." The Cronbach's Alpha Coefficient is a correlation coefficient that determines the correlation between the duplicate questions, thereby giving an indication of whether the participants have the same understanding of a question when it is asked in a slightly different way, more than once. In many cases, the opposite, inverted form of the same question can be asked later in the survey. An example of two duplicate questions would be one question early on in the survey such as "Did you find the web design effective to complete the tasks?" and later, toward the end of the survey, "Was the web design effective to complete the tasks?" Alternatively, a question asking the same issue of web effectiveness in a reversed manner can also be posed later in the survey in an inverted question such as "Did you find the web design ineffective to complete the tasks?"

In general, one or two duplicate question pairs are put into surveys of moderate size, up to 40 questions. It may be more helpful to insert more than one pair of duplicate questions into surveys that contain more than 40 questions. Also, if the survey has sections (for example, in a survey measuring web usability, sections may include usability of colors, layout, navigation, etc.) it is recommended to have one duplicate pair of questions for each of the sections to have freedom about determining the individual internal reliabilities of each section.

The Cronbach's Alpha coefficient is a correlation coefficient that produces a value between zero and one. The correlation between the duplicate questions is measured, and if the coefficient is equal to or greater than 0.7, then a survey is accepted as having high internal reliability (Cronbach, 1990). A set of duplicate questions, another set of duplicate, inverted questions, and a sample Cronbach's Alpha Coefficient computer output from the SAS (Statistical Analysis Software) computer package are presented in Table 58.4.

Internal reliability cannot be measured for surveys that contain open-ended questions. In empirical HCI research, however, most surveys with quantitative parts are required to have a satisfactory internal reliability coefficient in order to prove the reliability and validity of their results. Simply put, lack of internal reliability may result in the questions not measuring what they are intending to measure. Therefore, it is imperative in HCI research to insert at least one pair of duplicate questions into quantitative surveys.

More controversial than the internal reliability measure, the inter-rater reliability of a quantitative survey is concerned about the consistency among responses given by different participants to the same question. One argument is that in objective surveys, a consistency should be expected to some level among participant responses given to the same question in the same survey. While this argument may stay true to some extent in evaluation surveys, opinion surveys, as the name indicates, are about participants' opinions, which will obviously differ from person to person. Hence, it is arguable that the inter-rater reliability coefficient is a valid measure in opinion surveys. Additionally, a certain amount of variability is always present among responses to evaluation survey questions, even if the participants are all exposed to the exact same environment prior to the implementation of the surveys. The inter-rater reliability coefficient is a cor-

TABLE 58.4. Sample Duplicate Questions and Cronbach's Alpha Internal Reliability Coefficient SAS Computer Package Output

**Duplicate Pair with Same Question:**

1. How would you rate the convenience of this interface?

| ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|
| Very Inconvenient | Inconvenient | Neutral | Convenient | Very Convenient |

2. In general, how convenient was this interface for you to use?

| ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|
| Very Inconvenient | Inconvenient | Neutral | Convenient | Very Convenient |

**Duplicate Pair with Inverted Question (Inversion of the Scale Needed for Second Question):**

1. How easy were the tasks?

| ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|
| Very Difficult | Difficult | Not Difficult | Easy | Very Easy |

2. How difficult were the tasks?

| ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|
| Very Easy | Easy | Not Difficult | Difficult | Very Difficult |

**Sample SAS Output:**

| Variables | Alpha |
|---|---|
| Raw | 0.881884 |
| Standardized | 0.882042 |

Cronbach Coefficient Alpha with Deleted Variable

| | Raw Variables | | Standardized Variables | |
|---|---|---|---|---|
| Deleted | Correlation | | Correlation | |
| Variable | with Total | Alpha | with Total | Alpha |
| Question 1 | 0.717499 | . | 0.717499 | . |
| Question 2 | 0.717499 | . | 0.717499 | . |

Pearson Correlation Coefficients, N = 272

Prob > |r| under H0: Rho = 0

| | Question 1 | Question 2 |
|---|---|---|
| Question 1 | 1.00000 | 0.71750 |
| | | <.0001 |
| Question 2 | 0.71750 | 1.00000 |
| | <.0001 | |

relation coefficient among these responses given to the same question by different participants. And expecting a correlation as high as 0.7 among the participant responses may in most cases be not very realistic as a proof of reliability of a survey. Hence, while the inter-rater reliability coefficient is used in a number of survey types primarily in psychology, it is not seen as an essential measurement coefficient for survey reliability in relation to HCI research (Ozok & Salvendy, 2000). The inter-rater reliability coefficient is also used to determined how professionals in psychology and sociology rate the same situation, but this type is not covered in detail due to its lack of relevance to HCI.

Measuring the designed survey's reliability is crucial to producing useful results in HCI research. Therefore, comprehensive survey designers must pay attention to these reliability measures while designing their surveys, ensuring that the results obtained from their surveys in current and future studies will have high reliability, thereby improving the theoretical and practical impact of their research.

***Other survey evaluation issues.*** A large part of evaluation of the surveys usually happens after they are administered. However, in most cases an equally crucial evaluation of a survey happens just prior to the implementation. This evaluation is the actual *pilot testing* of the survey, sometimes also referred to as "pre-testing" (Dillman, 2000). As is the convention with most experimentation techniques in HCI research, the near-complete surveys can be administered to a group of participants. After this initial administration, the participants can be asked some questions about positive, negative, and missing issues in

the survey, and any questions or sections that were incomprehensible or unclear. Based on the feedback, the administrators can revise the survey and prepare it for the final administration. While there are no firm guidelines regarding the number of participants the pilot testing of surveys should be run on, in most cases a minimum of three participants is recommended for moderate-size surveys (less than 200 participants). For large-size surveys, up to 10 participants are generally useful (Dillman, 2000), although there is no upper limit for the number of participants to be used for the pilot study. Most surveys require revision after the pilot study, as in most cases there are some points the survey designers miss without the perspective of actual survey participants. In rare cases when no revisions are made to the surveys after the pilot survey administration, data obtained from the participants in the pilot can be included in the actual participant data pool.

How well a survey is designed is directly related to the validity and reliability of the results the research produces. Hence, the evaluation techniques covered in this section are crucial to the overall success of the designed survey and the research itself.

#### *Survey Design, Redesign, and Revision*

Design, redesign, and revision procedures for surveys to some extent bear some similarities to product design, redesign, and revision procedures. The initial design of surveys, as explained earlier, consists of generating questions based on the literature, expert opinion, and heuristics. Redesign and revision procedures mostly rely on the implementation of the survey on the entire group or a subgroup of participants. In most cases after a pilot test, surveys need revision which consists of changing or revising questions, scales, or instructions in the survey to make them clearer and more understandable for the participants. If there are errors in question structures or spelling errors, those are also located and eliminated after the pilot study. In rare cases, the required changes may be significant to the level that the survey may need redesign through the revision and change of most questions, scales, and instructions. It can be said that most small-scale survey revisions happen based on the feedback obtained from the pilot study.

Showing a certain amount of similarity to consumer products, frequently used surveys also need redesign and revisions over longer periods. Specifically in HCI, user habits and evaluation criteria for technology and technology-related products and issues change. It is therefore recommended that validated surveys that are used as empirical measurement tools in HCI should be reevaluated and updated about once a year to ensure they are up-to-date measurement tools and contain the latest additions to the HCI area as far as evaluation and opinion elements and techniques are concerned.

#### *Illustrative Survey Design Example*

Figure 58.3 presents a sample of a complete, generic paper-and-pencil survey in the example of a postexperimental task satisfaction survey. In the design of a survey measuring the Tablet PC usability issues among academic professionals, the first step is to develop a literature portfolio. This portfolio should cover literature on both mobile computer usability and Tablet PC usability. Next, researchers may send an inquiry to a group of academicians who use Tablet PCs, inquiring about major usability categories in relation to Tablet PCs in open-ended questions. Based on the input from literature and expert opinions, the researchers create an initial set of questions and scales, pilot-test it, and administer the survey, most likely in an environment where they give specific Tablet PC tasks to participants in a controlled environment, preceded by the actual survey administration.

#### *Challenges in Survey Design*

Survey design challenges mostly deal with possible mistakes in producing the survey questions and scales. Additionally, some problems may occur due to the questions having no validity backing. Therefore, the key for HCI survey researchers is to gain strong background in the area through literature and expert opinion before designing the surveys. After this background is gained, researchers are likely to have no difficulty designing the surveys with the appropriate number and content of questions and scales, and a comprehensive set of instructions to be presented to the participants on how they should fill out the survey.

## SURVEY IMPLEMENTATION IN HCI

Survey implementation can be categorized into open and controlled survey implementation environments. In this section, these two environments are first discussed, followed by a discussion of sample representativeness issues, an implementation example, and discussions of implementation challenges and emerging and future trends in survey design and implementation.

### Open vs. Controlled Implementation Environments

Survey implementation (also referred to as "survey administration") occurs in two alternative environments. In controlled survey implementation environments, participants fill out the survey in an environment specifically designated for and arranged according to their activity. An open implementation environment, on the other hand, does not contain any specific environmental traits for the participant to implement the survey. Open environments also mostly do not include any restrictions in time or other factors.

Controlled environments for implementation of surveys usually have the goal of preventing any distraction for the participant to hinder his or her understanding or judgment. Controlled survey environments are in most cases well-lit experimental rooms with appropriate equipment to make the participant moderately comfortable (often consisting of a chair and a table). Both computer and paper-and-pencil-based surveys can be implemented in either open or controlled environments. If the survey is implemented in a controlled environment in front of a computer, the survey implementers need to make sure that the computer's alignments (screen brightness, glare, screen distance, keyboard

**GENERAL SATISFACTION SURVEY**

**Please indicate how you personally feel about performing the different tasks.**
**Each of the statements below is something that a person might say about performing a job like the searching task through an interface. You are to indicate your own personal feelings about your work experience with the different tasks you just completed by marking the number most closely describes how much you agree with each of the statements.**

*1. My opinion of myself went up when I performed the tasks correctly.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*2. Generally speaking, I am very satisfied with performing the tasks.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*3. The tasks I performed were very meaningful to me.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*4. I felt the current web page structure design is good enough for me to perform the tasks.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*5. The tasks were usually interesting enough to keep me from getting bored.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*6. My own feelings were not affected much one way or the other by how well I performed the different tasks.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*7. I am in general satisfied with the kind of work I performed in the different web-page tasks.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*8. Most of the things I had to do to perform the tasks seemed useless or trivial.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*9. I felt uncomfortable when I performed the tasks incorrectly.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*10. I felt very satisfied with the accomplishment I got from performing the tasks.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*11. I felt very satisfied with the amount of independent thought and action I could exercise in the tasks.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*12. With current web page structure design, I felt difficult to perform the tasks efficiently and effectively.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*13. I felt very satisfied with the amount of challenge in these tasks.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

*14. I felt very satisfied with the level of mental effort required to perform the tasks.*

| ○ | ○ | ○ | ○ | ○ | ○ | ○ |
|---|---|---|---|---|---|---|
| Strongly Disagree | Disagree | Moderately Disagree | Neutral | Moderately Agree | Agree | Strongly Agree |

FIGURE 58.3. Sample of a paper-and-pencil satisfaction survey.

height, and other ergonomics issues) are optimized for the participant. In controlled environments, in most cases a survey administrator is also available to answer possible questions from the participant. These types of controlled survey administration environments are usually used to implement evaluation surveys as in most cases participants had just completed computer-based tasks and for them to be able to evaluate the interfaces or any other HCI-related environments, products, or procedures, controlled environments force them to do those evaluations immediately, while the memories of the items for them to evaluate are still fresh in their minds. Opinion surveys are generally not implemented in controlled environments. In some rare cases in which recording participant behavior during survey implementation is part of the experimentation, a controlled environment

can provide the equipment to nonintrusively record participant behavior/activities during the implementation of the survey.

While surveys are in some cases implemented in closed environments, doing so may be costly and time-consuming. Therefore, unless there is explicit need for a controlled environment, surveys are more commonly implemented in open environments. Open environments are environments of the participants' choosing, in most cases environments from their daily life. In open environments, a survey administrator is not present. Implementing surveys in open environments has the advantage of giving the participants the choice to choose the time and place of the implementation. This flexibility above all increases the ease of finding participants. Additionally, the freedom for the participants to fill out the survey at their convenience also improves their feeling of freedom and may increase their enthusiasm, thereby improving the accuracy of the survey results (Dillman, 2000). On the other hand, the surveys being filled out without the presence of a survey administrator will prevent the participants from asking any questions regarding the survey during the administration. Additionally, previous research has indicated that if participants have no particular motivation to fill out the survey, they may complete it very quickly without paying much attention, resulting in inaccurate survey results (Cochran, 1977).

Both controlled and open survey implementation environments have advantages and disadvantages. In most cases, however, open environments are faster and more convenient to collect the needed data due to the flexibility they offer to both the participants and implementers.

## Sample Representativeness in Survey Implementation

To ensure the validity of the results obtained from surveys, it is imperative to choose a representative sample of the target population to successfully implement the survey. It is common knowledge that the validity of the survey results improves with larger sample sizes. Therefore, researchers need to carefully choose both the sample sizes and the sample participants.

There are no strict rules for determining sample sizes in survey implementation. The size of the survey sample, meaning how many participants should fill out the survey, depends on the type of the survey and survey questions, as well as the number of questions in the survey. Thiemann and Kraemer (1987) summarized the statistical methods of determining sample sizes based on the number of variables being measured in the experiment. In most cases, surveys measuring general topics (for example, surveys about cell-phone-use habits of a certain population, such as college students) should be implemented on relatively large sample sizes, possibly no less than 60 participants for a survey consisting of up to 30 questions. For survey implementation, as a rule of thumb the number of participants should always be bigger than the number of questions in the survey. While there are no set rules on choosing sample sizes for survey implementation, large sample sizes always improve the probability of obtaining high validity of surveys. For more on how to calculate optimal sample sizes in survey implementation to obtain satisfactory statistical power, see Thiemann and Kraemer (1987).

When surveys are administered, one of the most critical issues is to administer the survey on a balanced, nonbiased, and representative sample. In general, for surveys administered online, the survey should be sent to an as big a potential participant pool as possible to ensure heterogeneity. A large group of potential participants should also be sought if the surveys are paper-and-pencil-based, by, for example, mailing out a large number of paper surveys. A balanced sample size in terms of race and gender ensures higher validity as well as a broader application of the results, and this can be achieved by sending the survey to a large base of participants. The demographic information collected also helps determine exactly whether the participant sample accurately represents the target population. As in any experimental or survey-based research, heterogeneity of the sample size allows the researchers to strengthen their argument that the results of their study are applicable and generalizable to the majority of their target population. For example, if there is a vast majority of males compared to females in the sample while the gender distribution in the actual target population is estimated to be about even, then the validity of the results may be argued upon as the variation of the results due to the females would not be taken into consideration in the population, and consequently in the conclusions derived from the survey results. Therefore, survey participant pools should be chosen carefully, and they should also adequately represent the target population. Especially in large sample sizes, sample characteristics may vary greatly, especially concerning age, education, and occupation demographics. In these cases, the demographics should be presented in detail as part of research results. In most cases, a detailed explanation of demographics concerning participants' education and occupation can justify the representativeness issue of survey results, as long as the levels of these attributes do not differ very greatly between the sample and estimated target population. In those cases, the results of the survey findings should indicate that the findings are likely to apply to the particular segment of the population that had an overwhelming majority among participants in the sample size. For example, suppose a survey on habits of general population concerning the use of cell phones has been conducted. If the vast majority of the participants (more than about two thirds) are university students, then the researchers should indicate in their report of results that they measured the cell phone habits of the university student population, which constitutes to a large percentage of avid cell-phone users (Ozok & Wei, 2004).

Survey participant pools require caution when they are chosen, and in cases when the researchers are convinced that the sample is not highly representative of the target population, they need to make clear that the results of their survey may possibly have a narrower focus. In most cases, such narrowing of the target population which the research is aimed at does not result in validity problems, but rather makes clear which population or population segment the results of the survey study apply to.

## Challenges in Survey Implementation

As indicated in previous sections, surveys are a relatively easy way of collecting data. They usually don't require expensive equipment for implementation, and with the advent of the Internet, can be easily distributed, filled out, returned, and analyzed. However, there are still some serious challenges in survey implementation in both paper-and-pencil and electronic environments.

Looking at the big picture, surveys are sometimes referred to as "subjective measurements." While the author of this chapter strongly disagrees with this statement, the distinction should be made between measurement environments where performance measurements are taken objectively and unequivocally through camera recordings and software, and environments where participants are asked to indicate what is going through their minds. In the latter environment, obviously, there is no way to ensure that participants are putting on the surveys exactly what they think about an issue, an environment, a tool, a product, or a procedure. It is not uncommon that participants fill out a survey without paying much attention, or even randomly mark responses without reading the questions. The survey-reliability measurement techniques to some extent prevent this type of random data from being used. For example, whether participants took the survey questions seriously can be determined by looking for discrepancies between Cronbach's Alpha duplicate questions. Additionally, strict instructions given to the participants at the beginning of the survey in written or spoken format can also to some extent improve the probability of participants taking the time to read the questions and give replies carefully. Researchers using surveys as their primary research tool always reserve the right to eliminate participant data that look ill fated or incomplete. However, the researchers need to have evidence in their hands that the participant did not complete the survey by obeying the rules that were presented to them, not on any other ground such as the participant responses not being in accordance with the majority of other participants or with the direction of results that are expected from the research.

Another challenge is the "return rate" of surveys. Response and return rates among surveys that do not offer any compensation is extremely low—less than 20% (Dillman, 2000). HCI research may involve lengthy surveys as well (more than 30 questions) which can potentially result in even lower return rates. Therefore, it is recommended that some sort of compensation should be offered to participants in survey research, whatever the resources will allow. This compensation may be small gifts (e.g., a pen or notepad), gift certificates, or cash compensation. Any of these types of incentives will surely improve the return rates of surveys (Miller, 1996). Additionally, surveys can contain statements to convince participants that they will also learn important issues concerning the research while filling out the surveys. It should be noted, however, that the practical issue of finding subjects should not bias the sample, and recruiting of participants should be arranged according to the data-collection needs of the research, not according to what kind of participant groups are the most practical to collect data from. Recruitment activities need to be targeted to ensure a representative sample.

In most implementation activities, participants should be given their privacy while filling out the survey, ensured that their data will be kept confidential, and be provided a comfortable environment. Otherwise, they may want to either quit or finish as soon as possible without any consideration of the accuracy of responses. It should also be noted that surveys are voluntary, and therefore survey implementers should indicate the voluntary nature of the surveys and not pressure the participants. Research indicates that when participants are pressurized to give accurate responses or when a mutual trust between the administrator and the participant is not established regarding the sincerity of both sides, they mostly produce very unreliable survey results (Dillman, 2000).

Other survey implementation challenges involve participants' interaction with the survey interface. In paper-and-pencil surveys, the fonts on the paper should have enough size and familiarity for all participants, a pencil with an eraser head will allow participants to correct their responses, and survey elements such as questions and scales should be adequately distinct from each other to prevent any mistakes. Survey designers should use a very simple language in the surveys and avoid any little-known words and sentence structures (Gendall, 1998). Surveys consisting of multiple pages should be clearly numbered on each page. This kind of a convenient interaction environment will improve participant enthusiasm and increase response rates as well as accuracy of survey results. Additionally, while some studies expressed concern about differences in survey responses among computer and paper-and-pencil surveys (Sanchez, 1992; Cole, 2005) and issues concerning the format in which online surveys are presented (Couper, Traugott, & Lamias, 2001; Couper, Traugott, Conrad, & Crawford, 2004; Kiernan, Kiernan, Oyler, & Gilles, 2005), a recent study indicated that the accuracy of survey responses did not significantly differ between surveys administered online and those administered paper-and-pencil (Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003).

In electronic surveys, the computer interface should be very simple and participants with little experience with computers should not have any difficulty using the interfaces. In most cases, participants are required to mark their responses with a mouse and type their responses on clearly marked, designated text spaces that have adequate size for visibility purposes. Some special equipment may offer some additional conveniences, such as touch screens. Screen glare and font sizes should be given consideration to. It can be recommended that computer surveys should be implemented on screens no smaller than 12 inches of diagonal size, with a refresh rate of at least 60 MHz. For surveys longer than one screen, scrolling should be minimized. It is recommended that each survey screen should not require more than two screen-heights of scrolling, and should be connected with hyperlinks or screen buttons, meaning once a participant completed a screen, he or she should be required to move on to the next screen by clicking on a screen button or a link. Besides the scrolling issue, if a participant sees a lengthy survey all presented on one screen, he or she may get discouraged to fill out the survey due to its length.

An additional potentially problematic item is the number of questions to ask participants in a survey. In most cases, the attention span of participants is very short, and surveys that do not offer any compensation are recommended to be shorter than 30 questions. In most cases, participants are not interested in spending more than 15 minutes in filling out surveys for which they don't get any compensation. There is always a tradeoff between the size of surveys, meaning the ability to collect all the necessary data, and the ability to recruit subjects. Long surveys are more difficult to recruit participants for. Researchers should think carefully about compensation methods (money, gifts, gift certificates) if they intend to implement large-scale surveys.

Finally, Internet surveys also carry the potential of technical difficulties due to the variety of computers the participants may be using. Schleyer and Forrest (2000) identified usability problems,

programming errors, and incompatibilities/technical problems as main problems identified in web-administered surveys. Therefore, Internet-based, especially web-based surveys should not require any scripts or plug-ins to run and if possible should consist of simple Hypertext Markup Language (HTML) code, as HTML is universally recognized by all browsers.

These are the major challenges currently the survey implementers have to deal with. However, with careful design and implementation, as well as strict instructions containing comprehensive information presented to the participants regarding the survey, the challenges can be easily overcome, resulting in valid and reliable survey results.

## EMERGING AND FUTURE TRENDS IN SURVEY DESIGN AND IMPLEMENTATION IN HCI

Surveys have retained their structure, more or less, for many decades. It can be said, however, that electronic and especially Internet-based implementation has changed the convenience level of survey implementation and data analysis in a significantly positive way. It is difficult to predict whether any new groundbreaking techniques will cause further leaps in survey design, development, and analysis, but if significant new developments will happen in the near future, they are likely to happen in the implementation technology. Internet surveys are on the rise, with the percentage of Internet-based surveys being on the rise for the past five years. While Internet-based surveys comprised 15% of all surveys implemented in 1999, this number increased to 70% in 2004, according to Nua Internet Surveys (2005). With the improvement in voice recognition and voice synthesis technologies, future surveys may eliminate the need of a visual interface; however, human-factors issues in these types of interfaces specifically for surveys are still to be explored. It is apparent that the number of surveys implemented through the Internet and other networks will continue to climb in the years to come, due to cost savings and a number of other convenience issues. In the future, HCI research is also likely to continue to use surveys as a main data-collection tool. With HCI research becoming a more integral part of technology design and creation (for example, human-factors engineers and software engineers working collaboratively in the software-design process), user surveys may become more integrated, collecting data for current or future users regarding both HCI and other technology issues. Additionally, computer literacy is increasing at a fast pace (U.S. Census Bureau, 2005), which allows HCI survey researchers more freedom in asking more sophisticated questions concerning interface evaluation, current user trends, and more. All that said, the contribution of surveys to HCI research is highly significant these days, and will likely continue to be so in the many years to come.

## References


Aiken, Lewis R., & Lewis A. (1996). *Rating scales and checklists: Evaluating behavior, personality, and attitudes* (1st ed.). NY: John Wiley & Sons.

Badre, A. (2002). *Shaping web usability: Interaction design in context*. Boston, MA: Addison Wesley Professional.

Baecker, R., Grudin, J., Buxton, W., & Greenberg, S. (1995). A historical and intellectual perspective. In Baecker, R. M., Grudin, J., Buxton, W. A. S., & Greenberg, S. (Eds.), *Readings in human-computer interaction: Toward the year 2000* (2nd ed.) (pp. 35–47). San Francisco, CA: Morgan Kaufmann Publishers, Inc.

Baker, R. (1998). The CASIC future. In M. P. Couper (Ed.), *Computer assisted survey information collection* (1st ed.) (pp. 583–604). New York, NY: John Wiley & Sons.

Belson, W. (1981). *The design and understanding of survey questions*. Aldershot, England: Gower.

Card, S. (1996). Pioneers and settlers: Methods used in successful user interface design. In M. Rudisill, et al. (Eds.), *Human-computer interface design: Success stories, emerging methods, and real-world context* (pp. 122–169). San Francisco, CA: Morgan Kaufmann Publishers.

Carini, R. M., Hayek, J. H., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College student responses to web and paper surveys: Does mode matter? *Research in Higher Education, 44*, 1–19.

Cochran, W. (1977). *Sampling techniques* (3rd ed.). New York, NY: John Wiley & Sons.

Cole, S. T. (2005). Comparing mail and web-based survey distribution methods: Results of surveys to leisure travel retailers. *Journal of Travel Research, 43*(4), 422–430.

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage Publications, Inc.

Couper, M. P., Traugott, M. W., Conrad, & Crawford (2004). What they see is what we get:Response options for web surveys. *Social Science Computer Review, 22*(1), 111–127.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly, 65*, 230–253.

Cronbach, L. J. (1990). *Essentials of psychological testing*. New York, NY: Harper & Row Publishing.

Dillman, D. (1991). The design and administration of mail surveys. *Annual Review of Sociology, 17*, 225–249.

Dillman, D. (1999): Mail and other self-administered surveys in the 21st century. *The Gallup Research Journal, Winter/Spring 1999*, 121–140.

Dillman, D. (2000). *Mail and Internet surveys: The tailored design method.* New York, NY: John Wiley and Sons, Inc.

Gendall, P. (1998). A framework for questionnaire design: Labaw revisited. *Marketing Bulletin, 9*, 28–39.

Hackman, R. J., & Oldham, G. R. (1980). Development of job diagnostic survey. *Journal of Applied Psychology, 60*(2), 159–170.

Jenkins, C. R., & Dillman, D. A. (1997). Towards a theory of self-administered questionnaire design. In L. Lyberg, P. Biemer, M. Collins, E. deLeeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 165–196). New York, NY: John Wiley & Sons.

Kiernan, N. E., Kiernan, M., Oyler, M. A., & Gilles, C. (2005). Is a web survey as effective as a mail survey? A field experiment among computer users. *American Journal of Evaluation, 26*(2), 245–252.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537–567.

Litwin, M. (1995). *How to measure survey reliability and validity.* Thousand Oaks, CA: Sage Publications.

Lund, E., & Gram, I. T. (1998). Response rate according to title and length of questionnaire. *Scandinavian Journal of Public Health, 26*(2), 154–160.

Medsker, G. J., & Campion, M. A. (1997). Job and team design. In G. Salvendy (Ed), *Handbook of Human Factors and Ergonomics* (3rd ed.) (pp. 450–489). New York: Wiley.

Miller, K. (1996). *The influence of difference techniques on response rates and nonresponse error in mail surveys.* Unpublished Master's thesis, Western Washington University, Ballingham, WA.

Myers, B. (1996). *A quick history of human computer interaction.* Carnegie Mellon University School of Computer Science Technical Report. CMU- S-96-163 and Human Computer Interaction Institute Technical Report. CMU-HCII-96-103, August, 1996.

Myers, B., Hollan, J., & Cruz, I. (1996). Strategic directions in human computer interaction. *ACM Computing Surveys, 28*(4), 794–809.

Nielsen, J. (1989). Coordinating user interface for consistency. *SIGCHI Bulletin, 20*, 63–65.

Nielsen, J. (1993). *Usability engineering.* London, UK: Academic Press.

Nua Internet Surveys (2005): nua.com, site accessed 9.15.2005.

Ozok, A. A., & Salvendy, G. (2000). Measuring consistency of web page design and its effects on performance and satisfaction. *Ergonomics, 43*(4), 443–460.

Ozok, A. A., & Salvendy, G. (2001). How consistent is your web design? *Behaviour and Information Technology, 20*(6), 433–447.

Ozok, A. A., & Wei, J. (2004). User perspectives of mobile and electronic commerce with a usability emphasis. *Proceedings of the ISOneWorld 2004 Conference*, Las Vegas, NE, Article 71.

Sanchez, M. E. (1992). Effect of questionnaire design on the quality of survey data. *Public Opinion Quarterly, 56*, 206–217.

Schleyer, T. K. L., & Forrest, J. (2000). Methods for the design and administration of web-based surveys. *Journal of American Medical Informatics Association, 7*(4), 416–425.

Shneiderman, B. (1992). *Designing the user interface: Strategies for effective human-computer interaction.* New York: Addison-Wesley.

Stone, D. H. (1993). Design a questionnaire. *British Medical Journal, 307*(6914), 1264–1266.

Thiemann, S., & Kraemer, H. C. (1987). *How many subjects?: Statistical power analysis in research.* Newbury Park, CA: Sage Publishing.

U.S. Census Bureau (2005). Computer and Internet use in the United States: 2003. *Annual Report Special Studies*, 23–108.

Zhang, Y. (2000). Using the Internet for survey research: A case study. *Journal of American Society for Information Science, 51*(1), 57–68.