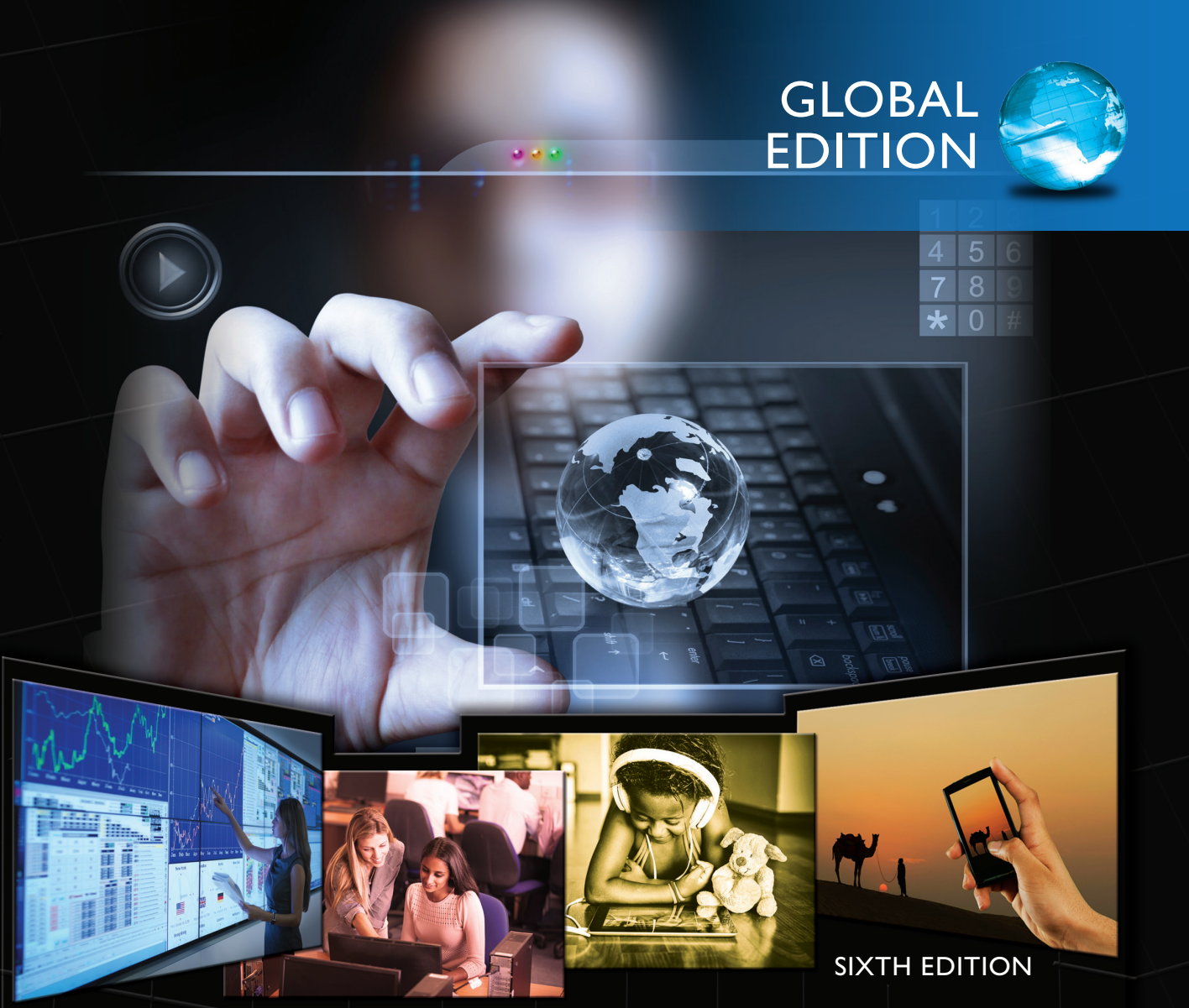
GLOBAL EDITION



SIXTH EDITION

Designing the User Interface

Strategies for Effective Human-Computer Interaction

Shneiderman • Plaisant • Cohen • Jacobs • Elmqvist

Pearson

DESIGNING THE USER INTERFACE

are developing guidelines for formative usability test reports, and some best practice guidelines are emerging. Key points are that it is important to understand the audience (who will be reading the report) and to keep the report concrete and specific.

5.4 Survey Instruments

User surveys (written or online) are familiar, inexpensive, and generally acceptable companions for usability tests and expert reviews. Managers and users can easily grasp the notion of surveys, and the typically large numbers of respondents (hundreds to thousands of users) confer a sense of authority compared to the potentially biased and highly variable results from small numbers of usability-test participants or expert reviewers. The keys to successful surveys are clear goals in advance and development of focused items that help to attain those goals. Two critical aspects of survey design are validity and reliability. Experienced surveyors know that care is needed during design, administration, and data analysis (Lazar et al., 2009; Cairns and Cox, 2011; Kohavi et al., 2013; Tullis and Albert, 2013). Additional information on surveys can be found in Chapter 4.

5.4.1 Preparing and designing survey questions

A survey form should be prepared, reviewed by colleagues, and tested with a small sample of users before a large-scale survey is conducted. Methods of statistical analysis (beyond means and standard deviations) and presentation (histograms, scatterplots, and so on) should also be developed before the final survey is distributed. In short, directed activities are more successful than unplanned statistics-gathering expeditions. Our experience is that directed activities also provide the most fertile frameworks for unanticipated discoveries. Since biased samples of respondents can produce erroneous results, survey planners need to build in methods to verify that respondents represent the population in terms of age, gender, experience, and other relevant characteristics.

It is important to pre-test or pilot-test any survey instrument prior to actual use. Users can be asked for their subjective impressions about specific aspects of the interface, such as the representation of:

- Task domain objects and actions
- Interface domain metaphors and action handles
- Syntax of inputs and design of screen displays

It may also be useful to ascertain certain characteristics about the users, including:

- Background *demographics* (age, gender, origins, native language, education, income)
- Experience with computers (specific applications or software packages, length of time, depth of knowledge, whether knowledge was acquired through formal training or self-teaching)
- Job responsibilities (decision-making influence, managerial roles, motivation)
- Personality style (introvert versus extrovert, risk taking versus risk averse, early versus late adopter, systematic versus opportunistic)
- Reasons for not using an interface (inadequate services, too complex, too slow, afraid)
- Familiarity with features (printing, macros, shortcuts, tutorials)
- Feelings after using an interface (confused versus clear, frustrated versus in control, bored versus excited)

Online and web-based surveys avoid the cost and effort of printing, distributing, and collecting paper forms. Many people prefer to answer a brief survey on a computer or other electronic device instead of filling in and returning a printed form, although there is a potential bias in the self-selected sample. Some surveys can have very large numbers of respondents. Some companies that provide computerized surveys include Survey Monkey, Survey Gizmo, Qualtrics, and Question Pro. Academic or educational discounts may be available.

Typically, participants are asked to respond to a series of statements according to the following commonly used *Likert* scale:

Strongly agree Agree Neutral Disagree Strongly disagree

The items in the survey could be similar to the following:

- I can effectively perform the task using this interface
- Items are placed where I expected to find them in the interface

Such a list of statements can help designers to identify problems users are having and to demonstrate improvements to the interface as changes are made; progress is demonstrated by improved scores on subsequent surveys.

Another approach is to use a set of bipolar semantically anchored items (pleasing versus irritating, simple versus complicated, concise versus redundant) that ask users to describe their reactions to using the interface. Users have to rate the items on 1-to-7 scales:

Hostile	1	2	3	4	5	6	7	Friendly
Easy to use	1	2	3	4	5	6	7	Difficult to use
Clear	1	2	3	4	5	6	7	Confusing

Yet another approach is to ask users to evaluate various aspects of the interface design, such as the readability of characters, use of the terminology, organization of the structure, or the meaning of the icons/controls. If users rate as poor one aspect of the system, the designers have a clear indication of what needs to be redone. If precise—as opposed to general—questions are used in surveys, there is a greater chance that the results will provide useful guidance for taking action.

Additional attention may be needed when dealing with special populations (see Chapter 2). For example, questionnaires for children must be in age-appropriate language, questionnaires for international users may need to be translated, larger fonts may be needed for older adults, and special accommodations may need to be made for users with disabilities.

5.4.2 Sample questionnaires

Questionnaires and surveys are commonly used in usability evaluation. Several instruments and scales have been developed and refined over time. The early questionnaires concentrated on elements such as clarity of fonts, appearance on the display, and keyboard configurations. Later questionnaires dealt with multimedia components, conferencing, and other current interface designs including consumer electronics and mobile devices. Here is some information on a few (most use a Likert-like scale):

The *Questionnaire for User Interaction Satisfaction* (QUIS). The QUIS (http://lap.umd.edu/quis/) has been applied in many projects with thousands of users, and new versions have been created that include items relating to website design. The University of Maryland's Office of Technology Commercialization licenses the QUIS. Special licensing terms may be available for students. Table 5.1 contains a portion of the QUIS, including an example for collecting computer experience data.

The *System Usability Scale* (SUS). Developed by John Brooke, it is sometimes referred to as the "quick and dirty" scale. The SUS consists of 10 statements

Examples of the specific satisfaction scale questions:					
5.4	Messages which appear on display:	confusing	clear	1 2 3 4 5 6 7 8 9	NA
5.4.1	Instructions for commands or choice:	confusing	clear	1 2 3 4 5 6 7 8 9	NA

TABLE 5.1

Questionnaire for User Interaction Satisfaction (QUIS) (© University of Maryland, 1997).

		Strongly disagree				Strongly agree
1	I think that I would like to use this system frequently	1	2	3	4	5
2	I found the system unnecessarily complex	1	2	3	4	5

TABLE 5.2
System Usability Scale (SUS) example (Brooke, 1996).

with which users rate their agreement (on a 5-point scale). Half of the questions are positively worded, and the other half are negatively worded. A score is computed that can be viewed as a percentage. Table 5.2 contains a sample from the SUS.

The *Computer System Usability Questionnaire* (CSUQ). A later development by IBM (based on the earlier PSSUQ) contains 19 statements to which participants respond using a 7-point scale. Table 5.3 contains a sample from the CSUQ.

The *Software Usability Measurement Inventory* (SUMI). Developed by the Human Factors Research Group (HFRG), it contains 50 items designed to measure users' perceptions of their affect (emotional response), efficiency, and control and of the learnability and helpfulness of the interface (Kirakowski and Corbett, 1993). Table 5.4 contains a sample from the SUMI.

The *Website Analysis and MeasureMent Inventory* (WAMMI) questionnaire. It was designed for web-based evaluations and is available in more than a dozen languages (http://www.wammi.com/).

Although many of these questionnaires were developed a while ago, they still serve as reliable and valid instruments. Some have been transformed by changing the focus of the items asked about. Specialized questionnaires have been developed and tested based on these proven instruments. One example is the Mobile Phone Usability Questionnaire (MPUQ), which consists of 72 items

			1 2 3 4 5 6 7		NA
1	Overall, I am satisfied with how easy it is to use this system.	Strongly disagree	• • • • • • •	Strongly agree	•
2	I can effectively complete my work using this system.	Strongly disagree	• • • • • • •	Strongly agree	•

TABLE 5.3
Computer System Usability Questionnaire (CSUQ) example.

		Agree	Undecided	Disagree
1	This software responds too slowly to inputs.	❑	❑	❑
2	I would recommend this software to my colleagues.	❑	❑	❑

TABLE 5.4

Software Usability Measurement Inventory (SUMI) example.

broken down into six factors: ease of learning and use, helpfulness and problem-solving capabilities, affective aspect and multimedia properties, commands and minimal memory load, control and efficiency, and typical tasks for mobile phones (Ryu, 2009). An example of sample questions from the MPUQ can be found in Table 5.5. The SUS has also been used with cellphones as well as interactive voice systems, web-based interfaces, and other interfaces and continues to be a robust and versatile tool. UMUX-LITE is another option as a shortened SUS (Lewis et al., 2013). As with any metric, any score should not be used in isolation. The best testing procedure, leading to the most confidence-inspiring results, would include triangulating the data from multiple methods, such as observations, interviews, logging of interface usage, and qualitative satisfaction data as well.

Writing and designing good questionnaires is an art as well as a science. Several books (Rubin and Chisnell, 2008; Sauro and Lewis, 2012; Tullis and Albert, 2013) and articles exist that provide further reading on use, validity, and development of good and valid questionnaires. In addition to the standard-type measures of satisfaction, specialized devices (e.g., mobile devices) and gaming interfaces may require unique measures such as pleasure, joy, affect, challenge, or realism. Some links to some older questionnaires include the following: Gary Perlman (http://garyperlman.com/quest/quest.cgi?form=USE) and Jurek Kirakowski (http://www.ucc.ie/hfrg/resources/qfaq1.html).

Examples of sample questions relating to mobile phones:
Is it easy to change the ringer signal?
Can you personalize the ringer signal with this product? If so, is that feature useful and enjoyable for you?
Do you feel excited when using this product?
Is it easy to use the phone book feature of this product?

TABLE 5.5

Mobile Phone Usability Questionnaire (MPUQ) example.

5.5 Acceptance Tests

For large implementation projects, the customer or manager usually sets objective and measurable goals for hardware and software performance. Many authors of requirements documents are even so bold as to specify the mean time between failures as well as the mean time to repair for hardware and, in some cases, software failures. More typically, a set of test cases is specified for the software, with possible response-time requirements for the hardware/software combination (see Chapter 12). If the completed product fails to meet these acceptance criteria, the system must be reworked until success is demonstrated.

These notions can be neatly extended to the human interface. Explicit acceptance criteria should be established when the requirements document is written or when a contract is offered. Rather than use the vague and misleading criterion of "user friendliness," measurable criteria for the user interface can be established for the following:

- Time for users to learn specific functions
- Speed of task performance
- Rate of errors by users
- User retention of commands over time
- Subjective user satisfaction

An acceptance test for a food-shopping website might specify the following:

> The participants will be 35 adults (25–45 years old), native speakers with no disabilities, hired from an employment agency. They will have moderate web-use experience: one to five hours/week for at least a year. They will be given a five-minute demonstration on the basic features. At least 30 of the 35 adults should be able to complete the benchmark tasks within 30 minutes.

Another testable requirement for the same interface might be this:

> Special participants in three categories will also be tested: (a) 10 older adults aged 55–65; (b) 10 adult users with varying motor, visual, and auditory disabilities; and (c) 10 adult users who are recent immigrants and use English as a second language.

Since the choice of the benchmark tasks is critical, pilot testing must be done to refine the materials and procedures used. A third item in the acceptance test plan might focus on retention:

> Ten participants will be recalled after one week and asked to carry out a new set of benchmark tasks. In 20 minutes, at least eight of the participants should be able to complete the tasks correctly.