

Chapter 13

AN EVALUATION FRAMEWORK

13.1 Introduction

13.2 DECIDE: A Framework to Guide Evaluation

Objectives

The main aims of this chapter are to:

- Introduce and explain the DECIDE framework.
- Discuss the conceptual, practical, and ethical issues involved in evaluation.

13.1 Introduction

As products evolve from initial ideas through conceptual design and prototypes, iterative cycles of design and evaluation help to ensure that they meet users' needs. However, deciding when and how to evaluate a product requires careful consideration and may be different for different kinds of products. This can be challenging given the rapid development of a wide variety of mobile and distributed systems.

The design process starts with the designers working to develop a product that meets users' requirements, but understanding requirements tends to happen by a process of negotiation between designers and users over a period of time. As designers understand users' needs better, their designs reflect this understanding. Similarly, as users see and experience design ideas, they are able to give better feedback that enables the designers to improve their designs. The process is cyclical, with evaluation facilitating understanding between designers and users.

Evaluation is driven by questions about how well the design or particular aspects of it satisfy requirements and offer appropriate user experiences. Some of these questions provide high-level goals to guide the evaluation. For example, does this product excite users so that they will buy and use it? Others are much more specific. Can users find a particular menu item? Do they interpret a particular graphic as the designers intended and do they find it attractive? Practical constraints play a big role in shaping how evaluation is done: tight schedules, low budgets, or limited access to users constrain what evaluators can do and may prompt them to seek new ways of doing evaluations, e.g. using crowdsourcing as discussed in Chapter 12. There are ethical considerations too: medical records are confidential, certain

areas of people's homes are private, and so is some of the information that users put on their Facebook and other social networking sites.

There is a broad repertoire of evaluation methods that can be tailored for specific circumstances, and experienced designers get to know what works and what doesn't. When planning evaluations, evaluators must consider the nature of each product, the target user population, where the product will be used and the contexts of use, as well as logistical issues. Planning evaluation studies involves asking questions about the process and anticipating potential problems. Within interaction design there are many books and websites that list different techniques and guidelines for conducting an evaluation, but there is very little overarching guidance for how to plan an evaluation. To help you, we propose the DECIDE framework, which provides a structure for planning evaluation studies.

13.2 DECIDE: A Framework to Guide Evaluation

Well-planned evaluations are driven by goals that aim to seek answers to questions, which may be stated explicitly, upfront, as in usability testing or may emerge as the evaluation progresses, as in ethnographic evaluation. The way questions are stated also varies depending on the stage of design when the evaluation occurs. Questions help to determine the kind of evaluation methods that are used. Practical issues need to be thought about, such as the cost, which is affected by the amount of time available to carry out the evaluation, the availability of participants, and the availability of suitable equipment. Ethical issues must also be considered, particularly when working directly with users, or with data that might be traceable back to users, e.g. in the collaborative game example discussed in Chapter 12. Evaluators must also have enough time and expertise to evaluate, analyze, interpret, and present the data that they collect so that it is meaningful to designers and to users. The DECIDE framework provides a checklist to help you plan your evaluation studies and to remind you about the issues you need to think about. It has the following six items:

1. Determine the goals
2. Explore the questions
3. Choose the evaluation methods
4. Identify the practical issues
5. Decide how to deal with the ethical issues
6. Evaluate, analyze, interpret, and present the data.

A list has the tendency to suggest an order in which things should be done. However, when working with the DECIDE framework, it is common to think about and deal with items iteratively, moving backwards and forwards between them after taking the first pass through each one. Making decisions about some items will impact others. For example, the goals of the evaluation will influence the questions and they in turn influence choice of methods, which are also guided by an underlying philosophy, e.g. ethnography or experimental design, which also guides data analysis. For example, data collected from laboratory-based experiments would not be analyzed using Grounded Theory. Each item in the DECIDE framework is related to the other items in various ways and so working iteratively is essential.

13.2.1 Determine the Goals

What are the high-level goals of the evaluation? Who wants it and why? An evaluation to help clarify that requirements have been met in an early design sketch has different goals from an evaluation to select the best representation of a metaphor for a conceptual design, or an evaluation to fine tune a website, or to examine how mobile technology changes working practices, or to inform how the next version of a product should be changed, or to explore the impact of ambient technology in a social space, or to investigate what makes collaborative computer games engaging.

Goals guide the evaluation by helping to determine its scope, so identifying what these goals are is the first step in planning an evaluation. For example, we can restate the first general goal statement mentioned above more clearly as: does the design sketch address the description of requirements?

ACTIVITY 13.1

What might an evaluation goal be for a new system aimed at teaching museum visitors about Greek jewelry.

Comment

A goal might be to find out whether visitors can discover more about the designers of the jewelry and its provenance when visiting the exhibit. ■

13.2.2 Explore the Questions

In order to make goals operational, we must clearly articulate the questions to be answered by the evaluation study. For example, a goal of finding out why some customers prefer to purchase paper airline tickets over the counter rather than e-tickets can be broken down into a number of relevant questions for investigation. What are customers' attitudes to e-tickets? Perhaps they don't trust the system and are not sure that they will actually get on the flight without a ticket in their hand. Maybe they think that thieves will steal credit card information and go on a spending spree with their card, or that the information that they type in will enable someone to steal their identity. Does the electronic system have a bad reputation? Perhaps a neighbor has complained that it took him over an hour to work out how to select a flight. Is the user interface to the ticketing system so poor that they can't use it? Maybe some people can't complete the transaction. Maybe some people like the social interaction with a ticketing agent.

Questions can be broken down into very specific sub-questions to make the evaluation even more fine-grained. For example, what does it mean to ask, 'Is the user interface poor?' The sub-questions that this question suggests might be as follows. Is the system difficult to navigate? Is the terminology confusing because it is inconsistent? Is the response time too slow? Is the feedback confusing or maybe insufficient? Sub-questions can, in turn, be further decomposed if even more specific issues need to be addressed.

ACTIVITY 13.2

The Hello.Wall was developed as an ambient display in which sensors and LEDs were embedded in a wall-like structure and was designed to blend in with the environment (Streitz *et al.*, 2005). It measured 1.8 meters wide by 2 meters high (see Figure 13.1). The aim was to enhance people's awareness of their environment by changing the pattern of LED lights in response to who was passing by or watching it. Different light patterns on the display correspond to different types of information. Some of the patterns created were public and others private to specific people.

The Hello.Wall was intended to have an aesthetic impact that would help to create a mood in the space that it occupies and to influence social gatherings. There are three different zones of interaction created around the Hello.Wall: (i) an interaction zone, which is close to the wall, (ii) a notification zone, which is further away, and (iii) an ambient zone, which is still further away. In each zone the wall generates interaction patterns that are appropriate. People's presence in the ambient zone contributes to the ambient patterns; when they move into the notification zone they are notified on a cell phone; and when in the interaction zone they can interact directly with the patterns on the Hello.Wall (see Figure 13.2).

Imagine you have been asked to evaluate the Hello.Wall. Your goal is to find out how people react to the Hello.Wall and whether they like it. Based on what you know about the Hello.Wall write four or five questions that you could investigate. ■

Comment

Possible questions include:

- Do users notice the HelloWall?
- For those who do notice it, how do they react to it?
- Do they understand how it works when they are at different distances from the wall?
- Do they seem to enjoy interacting with it?
- Do they tell others about it? If so, what do they say?

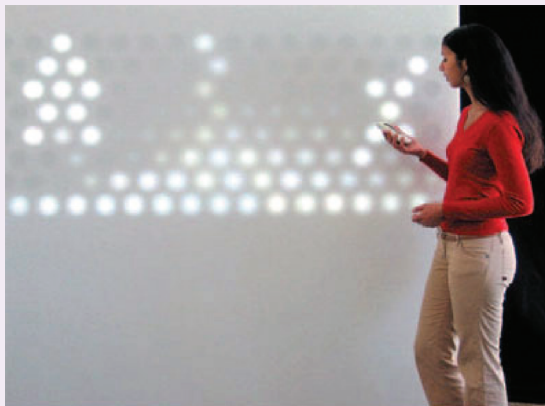


Figure 13.1 The Hello.Wall – an ambient display showing a continual display of high-quality aesthetic patterns

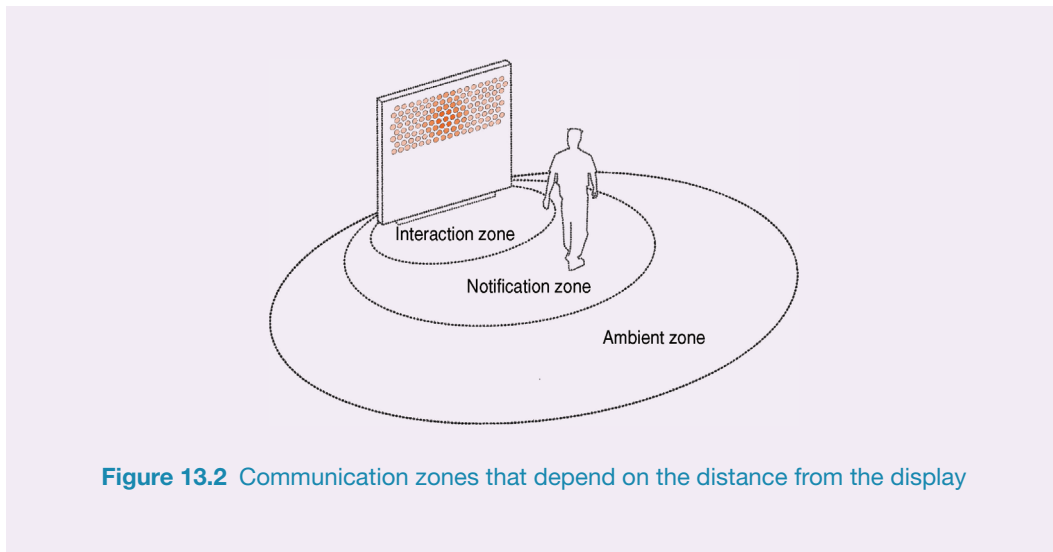


Figure 13.2 Communication zones that depend on the distance from the display

13.2.3 Choose the Evaluation Methods

Having identified the goals and articulated some questions, the next step is to choose the evaluation methods that you will use. Your choice will depend on what data is needed to answer the questions and which theories or frameworks are appropriate to your context. For example, you would choose a usability test if you wanted to evaluate a predetermined set of tasks for an interface – which you could not do with a field study. Conversely, it would be more appropriate to evaluate how an ambient display, such as the Hello.Wall, affects people’s spontaneous social interactions by carrying out an in the wild study – which would be difficult to do, if not impossible, in a laboratory setting. Practical issues may also constrain which method is selected. For example, the methods that seem most appropriate may be too expensive, or may take too long, or may require equipment or expertise that is not available, so compromises are needed.

Sometimes, combinations of methods are used as they enabled richer data to be gathered. Data collected using different methods gives a broad picture of how well the design meets the usability and user experience goals that were identified during the requirements activity.

ACTIVITY 13.3

1. Which evaluation methods would you choose to evaluate the Hello.Wall? How would you collect data?
2. What questions would you ask if you had chosen to use an interview?

(Continued)

Comment

1. A field study would be most appropriate for evaluating how people react and respond to the Hello.Wall in its natural environment. One way of collecting data would be to use observation, maybe by discretely positioning video cameras to record what people do when they are in the vicinity of the Hello.Wall. Interviews and questionnaires could also be used to collect data from people once they have moved out of its range.
2. Some possible questions are:
 - (a) What were your initial reactions when the Hello.Wall reacted to your presence?
 - (b) What did you actually do when you saw the Hello.Wall?
 - (c) How did it respond?
 - (d) How did that make you feel?

Probing for more information would be useful depending on how users answer these questions. ■

13.2.4 Identify the Practical Issues

There are often many practical issues to consider when conducting an evaluation, and it is helpful to know in advance what these are. However, even experienced evaluators encounter unexpected events, which is why it is useful to do a pilot study before the actual study (discussed in Chapter 7). This helps identify any problems that can be rectified, such as participants not understanding instructions or the placement of a display in situ being inappropriate (e.g. making it difficult to read because of direct sunlight). Issues that should be taken into account include access to appropriate participants, facilities, and equipment, whether schedules and budgets are realistic, and whether the evaluators have the appropriate expertise to conduct the study. Depending on the availability of resources, compromises may need to be made. For example, evaluators may wish to perform usability tests using 20 participants and then to run a three-week-long study in a natural setting, but the budget available for the study may only cover the cost of 10 testers and a shorter field study. It may be important to get feedback to designers quickly so time is traded off against the size and therefore the comprehensiveness of the study. When doing a field study of cell phone design it may be difficult to follow the users, as they are likely to be highly mobile. Furthermore, the cell users may go into places such as bathrooms, and bedrooms, where evaluators cannot go. Contingency plans are therefore needed to deal with such situations. There may be other surprises that require evaluators to take decisions on the spot. For example, it may not be possible to ride in the taxi or car with users because there isn't enough room. No evaluation is going to be perfect, and a good study doesn't require the evaluator to see how the product is used 100% of the time, but it is helpful to be aware of the kind of compromises that may be necessary.

Participants – Users, Potential Users, and Expert Evaluators

It goes without saying that a key aspect of an evaluation is involving appropriate participants. In the case of an inspection method and for evaluations based on models, these participants would be user experience consultants or experts. For methods involving users, the choice of

participant would focus on the characteristics of the anticipated user population and finding participants who represent the population of users for whom a product is targeted. This generally involves identifying people with a particular level of experience, e.g. novices or experts, or people with a range of expertise, as you saw in the e-skiing and crowdsourcing case studies in the previous chapter. For example, it would be no use asking people who never ski to participate in the e-skiing evaluation. The number of males and females within a particular age range, cultural diversity, educational experience, and personality differences may also need to be taken into account, depending on the kind of product being evaluated. Questionnaire surveys require large numbers of participants, so ways of identifying and reaching a representative sample of participants are needed. In field studies, sometimes it is necessary to observe what people do over a period of time and space. This will entail shadowing them, recording conversations, taking pictures, and so on. However, some people may not be comfortable with such a level of surveillance. For example, in a study of privacy practices among Facebook users, Mancini *et al* (2009) report that some of the volunteers who had previously come forward decided not to take part in the study as soon as they heard about the extent of the researchers' proposal to spend time with them. One candidate participant said she found the idea 'spooky,' while others said they thought it was 'weird' or 'strange.'

Another issue to consider is what the participants will be expected to do and for how long. The tasks used in a usability study should be representative of those for which the product is designed. However, there are no written rules about the length of time that a participant should be expected to spend on an evaluation task. Ten minutes is too short for most tasks and 2 hours is a long time; so what is reasonable? Task times will vary according to the type of evaluation, but when tasks go on for more than 20 minutes, consider offering breaks. It is accepted that people using desktop computers should stop, move around, and change their position regularly after every 20 minutes spent at the keyboard to avoid repetitive strain injury. Evaluators also need to put participants at ease so they are not anxious and will perform normally; it is important to treat them courteously. Participants should not be made to feel uncomfortable when they make mistakes. Greeting participants, explaining that it is the product that is being tested and not them helps to put participants at ease.

In studies in natural settings, the onus is on the evaluators to fit in with the participants and to cause as little disturbance to participants and their activities as possible. This requires practice, and even anthropologists who are trained in ethnographic methods may cause unforeseen changes (see the Dilemma box below).

DILEMMA

Is it possible to study people's behavior without influencing it?

A newspaper article reported on how an anthropology student traveling through northern Kenya happened by chance to come upon an unknown tribe. He studied their rituals and reported the study in his PhD dissertation and published several articles in acclaimed journals.

(Continued)

The study drew considerable attention because finding an unknown tribe is unusual in this day and age. It is the dream of many anthropologists because it allows them to study the tribe's customs before they are changed by outside influences. Of course, having published his work, the inevitable happened; more anthropologists made their way to the village and soon members of the tribe were drinking coke and wearing T-shirts from prestigious universities and well-known tourist destinations. The habits of these outsiders gradually changed the tribe's behavior.

Ethnographers face this dilemma: is it possible to study people's behavior without changing it in the process? ■

Facilities and Equipment

The recording equipment chosen for an evaluation needs to be considered in terms of how appropriate it is for collecting the data required. For example, when using video you need to think about how you will do the recording: how many cameras to use and where to put them. Some people are disturbed by having a camera pointed at them, and will not act normally, so you need to think about how you can make them feel comfortable. There may also be times when things happen very rapidly. For example, mobile device users move from place to place, sometimes with little or no warning, so how will you capture the data? You could take pictures, record comments, and write a few notes. These will be helpful reminders for you to write up a fuller account when the observation session has ended.

ACTIVITY 13.4

The evaluators in the e-Skiing study (Chapter 12) described some of the logistics that they needed to consider. What were they?

Comment

In the e-skiing case study it was tricky to get the various components of the system to work in a coordinated way as planned. They also needed to make sure that the physical size of the head-mounted components fitted on the skiers' helmets. There was also an unexpected finding – the skiers did not want to take time off the slopes or communicate from the slopes about their experiences using the system. Instead, they preferred to have a discussion in the bar after skiing. The evaluators agreed to this even though it meant changing their original evaluation plan. ■

Schedule and Budget Constraints

Planning evaluations that can be completed on schedule and within budget is important, particularly in commercial settings. Often compromises need to be made given the resources and time available.

Expertise

Different evaluation methods require different kinds of expertise. For example, running usability tests requires knowledge about how to control the testing conditions, and video recording. Performing an experiment requires knowledge of experimental design and usually statistics as well. If you need to analyze your results using statistical measures and you are unsure of how to do this, then consulting a statistician before starting the evaluation and then again during data collection and analysis can be a good strategy. Field studies in the wild require ethnographic experience. Before doing a pilot study and getting more deeply into the evaluation it is important for you to ask yourself: does the evaluation team have the expertise needed to do the evaluation being planned?

ACTIVITY 13.5

In the study to investigate the collaborative digital ice hockey game (in Chapter 12), the evaluators had to consider several practical issues. What were they?

Comment

The evaluators collected physiological data, e.g. heart and breathing rates, so they had to ensure that they did not cause physical or emotional harm to the participants. Expertise was needed to use the recording equipment which was strapped to the participants, so the study had to be done in a controlled setting. They also had to find participants whose ability to play the game was similar and who were willing to participate. ■

13.2.5 Decide How to Deal with the Ethical Issues

When collecting data during an evaluation it is necessary to consider ethical issues (see Chapter 7). The Association for Computing Machinery (ACM) and many other professional organizations provide ethical codes (Box 13.1) that they expect their members to uphold, particularly if their activities involve other human beings. In the case of the ACM they recognized as early as 1992 that it is important to protect humans who participate in studies either directly or indirectly through having their activities tracked and logged. Participants' privacy has to be protected, which means that their names should not be associated with data collected about them or disclosed in written reports (unless they give explicit permission). Personal records containing details about health, employment, education, financial status, and where participants live must be confidential. Similarly, it should not be possible to identify individuals from comments written in reports.

BOX 13.1**Association for Computing Machinery (ACM) Code of Ethics**

The ACM code outlines many ethical issues that professionals involved with designing and developing digital applications are likely to face. Section 1 outlines fundamental ethical considerations, while section 2 addresses additional, more specific considerations of professional conduct. Statements in section 3 pertain more specifically to individuals who have a leadership role. Principles involving compliance with the code are given in section 4. Three principles of particular relevance to this discussion are:

- Ensure that users and those who will be affected by a system have their needs clearly articulated during the assessment of requirements; later the system must be validated to meet requirements.
- Articulate and support policies that protect the dignity of users and others affected by a computing system.
- Honor confidentiality.

(For a broad discussion of the code, read Anderson *et al*, 1993.) ■

Most professional societies, universities, government, and other research offices require researchers to provide information about activities in which human participants will be involved. They do this to protect participants by ensuring that they are not endangered physically (e.g. in medical studies) or emotionally and that their right to privacy is protected. This documentation is reviewed by a panel and the researchers are notified whether their plan of work, particularly the details about how human participants and data collected about them will be treated, is acceptable. Drawing up such an agreement is mandatory in many universities and major organizations. Indeed, special review boards generally prescribe the format required and many provide a detailed form that must be completed. Once the details are accepted the review board checks periodically in order to oversee compliance. In American universities these are known as Institutional Review Boards (IRB). Other countries use different names and different forms for similar processes. Over the years IRB forms have become increasingly detailed, particularly now that much research involves the Internet and people's interaction via communication technologies. Several lawsuits at prominent universities have heightened attention to IRB compliance to the extent that it sometimes takes several months and multiple amendments to get IRB acceptance. IRB reviewers are not only interested in the more obvious issues of how participants will be treated and what they will be asked to do, they also want to know how the data will be analyzed and stored. For example, data about subjects must be coded and stored to prevent linking participants' names with that data. This means that names must be replaced by codes or pseudonyms that must be stored separately from the data and stored in a locked place.

ACTIVITY 13.6

An online gaming company wants to find out the impact of a new YouTube advertising video that shows a demo of a new online game for teenagers. Imagine the company hires you to conduct online interviews with 20 teens. What privacy issues might you need to consider?

Comment

You would need to discuss how you will perform the interviews, how you will collect the data, and how the data will be stored, analyzed, and reported. For each of these, you will need to specify privacy and security considerations. For example, each participant could be assigned a code or pseudonym by the evaluators. The codes or pseudonyms and the names to which they relate will be stored separately from the data. At no time will real names be used; nor will there be reference to any markers that could enable identity of the participant, e.g. where the participant lives or works, their gender or ethnicity, and whether these are distinguishing features among the pool of participants. ■

Each participant in an evaluation study may be required to sign an informed consent form. The example in Box 13.2 is for a study performed by a doctoral student, to evaluate how undergraduate students use their iPhones and iPod Touches for social interaction. These undergrads were part of a University of North Park initiative to provide them with these devices along with a suite of software developed to support their studies. Whether or not an informed consent form is needed in any one setting depends on the relationship between the evaluator and the participants; this is touched upon below and was discussed in Chapter 7. In studies involving controlled settings, it is more likely that such a consent form will be required.

BOX 13.2

An example of an Informed Consent Form

Project Title	Social Interaction Through Mobile Devices
Why is this research being done?	<i>This is a research project being conducted by Prof. Brown and Daniella Green at the University of North Park. We are inviting you to participate in this research project because you take part in the university's Mobility Initiative. The purpose of this research project is to understand how students use mobile devices as part of their social interactions.</i>

(Continued)

<p>What will I be asked to do?</p>	<p><i>The procedures involve participating in a focus-group interview that will discuss the use of mobile devices. The questions will address several topics, including:</i></p> <ul style="list-style-type: none"> • <i>What is “social interaction”</i> • <i>How do students use mobile devices to interact with each other on and off campus?</i> • <i>Which applications are preferred by students?</i> • <i>Does the use of mobile devices change students’ social interaction habits?</i> • <i>What other applications do you use for social interaction?</i> <p><i>After the focus-group interview you will also be asked to fill out a brief questionnaire. The total time for your participation will be around 2 hours. In order to accurately record your answers to the group-interview questions we would like to create a video recording of the interview.</i></p> <p><i>The study will take place at the University of North Park. Upon completion of your participation in the focus group, you will receive a \$15 coupon for the university book store.</i></p>
<p>What about confidentiality?</p>	<p><i>This research project involves making videotapes of the focus group, in order to accurately record the group’s answers to the questions that will be asked during the interview.</i></p> <p><i>We will do our best to keep your personal information confidential. Group-interview transcripts and questionnaires will only be shared with research team members. Quotes from group-interviews may be used in publications, but will not be attributable to individuals, since we will remove all personally identifiable information (e.g., names or uniquely identifying characteristics). Data will be stored on password-protected personal computers or servers, or if printed out, in a locked cupboard. Video recordings will also be kept in a locked cupboard. The data will be erased (or destroyed) when the project and resulting publications are complete.</i></p> <p><i>Please check one of the following:</i></p> <p><input type="checkbox"/> <i>I agree to be videotaped during my participation in this study.</i></p> <p><input type="checkbox"/> <i>I do not agree to be videotaped during my participation in this study.</i></p> <p><i>If we write a report or article about this research project, your identity will be protected to the maximum extent possible. By signing this consent form, you acknowledge that we can record and analyze the interview and use quotes in publications after they have been made anonymous.</i></p> <p><i>Your information may be shared with representatives of the University of North Park or governmental authorities if you or someone else is in danger or if we are required to do so by law. In accordance with legal requirements and/or professional standards, we will disclose to the appropriate individuals and/or authorities information that comes to our attention concerning child abuse or neglect or potential harm to you or others.</i></p>

What are the risks of this research?	<i>There are no known risks associated with participating in this research project.</i>	
What are the benefits of this research?	<i>This research is not designed to help you personally, but the results may help the investigator learn more about the ways students use mobile devices for social purposes. We hope that, in the future, other people might benefit from this study through improved understanding of these practices.</i>	
Do I have to be in this research? May I stop participating at any time?	<i>Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify. However, only participants who complete their participation in the focus group and answer the questionnaire will receive the \$15 coupon for the university bookstore.</i>	
What if I have questions?	<i>This research is being conducted by Daniella Green of the College of Information Studies at the University of North Park. If you have any questions about the research study itself, please contact Daniella Green, at: 4171 Elizabeth Bldg, University of North Park, tel. (601) 123 4567 or at daniella@unipark.edu. <i>If you have questions about your rights as a research subject or wish to report a research-related injury, please contact: Review Board Office, University of North Park, North Park; (e-mail); rb@unipark.edu (telephone) 601-555-5555</i> <i>This research has been reviewed according to the University of North Park review procedures for research involving human subjects.</i></i>	
Statement of Age of Subject and Consent	<i>Your signature indicates that: you are at least 18 years of age; the research has been explained to you; your questions have been fully answered; and you freely and voluntarily choose to participate in this research project.</i>	
Signature and Date	NAME OF RESEARCHER:	DATE:
	SIGNATURE:	

The following list will help ensure evaluations are done ethically and that adequate steps have been taken to protect users' rights:

- Tell people the goals of the study and exactly what they should expect if they agree to participate. The information given to them should include outlining the process, the approximate amount of time the study will take, the kind of data that will be collected, and how that data will be analyzed. The form of the final report should be described and, if possible, a copy offered to all the participants. Many evaluators and researchers also invite participants to

comment, particularly if the evaluation involves interviews, focus groups, and observations that could be interpreted differently depending on the experience and knowledge of the evaluators. In such circumstances sharing the report with participants serves two purposes. It is a way of checking that you have really understood what the participants were doing or saying. It is also respectful and a way of acknowledging your appreciation for their participation. Any payment offered should also be clearly stated.

- Be sure to explain that contact information and demographic, financial, health, or other sensitive information that participants disclose or is discovered during the evaluation is confidential. A coding system should be used to record each participant's data and, the code and the person's demographic details should be stored separately from the data. Anonymity should be promised.
- Make sure participants know that they are free to stop the evaluation at any time if they feel uncomfortable with the procedure.
- Consider your relationship with the participants and decide whether it is appropriate to provide incentives such as food, book tokens, or financial payment. For example, if it is your child taking part in a colleague's study, would a gift token or a toy be more appropriate than offering payment as an incentive? Some universities do not allow cash payments to students, or the opportunity for students to earn extra credit towards a class grade. Some do not allow any kind of reward, and some take a middle road; they will allow certain kinds of rewards, e.g. a book or music token. See Chapter 7 for further discussion.
- Ask participants for permission in advance to quote them, promise them anonymity, and offer to show them a copy of the report before it is distributed.
- Avoid including quotes or descriptions that inadvertently reveal a person's identity by using numbers or fictitious names to record and identify individuals. Where quotes are reported to illustrate findings then it is convention to replace words that would reveal the source with representative words in square brackets. For example, if the study was evaluating a university's information system and one of the participants commented 'When I tried to send a message to Harry Jones about my meeting with Mary Ann Green the whole system suddenly froze,' then the comment would be quoted as 'When I tried to send a message to [. . .] about my meeting with [. . .] the whole system suddenly froze.' Also avoid using descriptions that could identify a person. For example, if a focus group involves nine men and one woman, using the pronoun 'she' in the final report will clearly identify the one woman present.
- Whether you quote a participant or not you will gain that person's trust by offering to give them a draft of your final report and inviting them to give feedback.



"Mind filling out this
evaluation form first?"

ACTIVITY 13.7

1. Think back to the e-skiing and collaborative online game case studies in Chapter 12. What ethical issues might the evaluators have considered?
2. How might this be different for an evaluation of a medical system.

Comment

1. In the e-skiing study the evaluators probably considered whether the device endangered the skiers by getting in the way of their movements. Some skiers are very competitive so they may not have wanted other participants to see information about their performance. They would also want personal contact information to be treated confidentially. In the collaborative online game players may not want others to know their strategies. Some players might not want others to see the physiological data collected when they played the game. Similar to the e-skiing case study, they would also want personal contact information to be treated confidentially.
2. All the same ethical issues apply when evaluating medical systems but ensuring that personal data is kept confidential is even more important, particularly data about medical conditions that could be embarrassing or impact the participants rights in any way, e.g. to obtain medical insurance. ■

The explosion in Internet and web usage has resulted in more research on how people use technologies such as Facebook, Wikipedia, and Twitter, and on their effects on everyday life. Consequently, there are many projects in which developers and researchers log users' interactions, analyze blogs and microblogs, or examine conversations in chat rooms, discussion forums (Maloney-Krichmar and Preece, 2005), and on social networking sites. These studies can be done without users knowing that they are being studied. This raises ethical concerns, chief among which are issues of privacy, confidentiality, informed consent, and appropriation of others' personal stories (Bruckman, 2004). People often say things online that they would not say face to face. Furthermore, many people are unaware that the personal information they share online can be read by someone with technical know-how years later, even after they have deleted it from their personal mailbox (Erickson *et al*, 1999).

Ethical guidelines have been developed to protect online users but it takes time to develop guidelines and the pace of new research has tended to outpace development of robust ethical practices (Bruckman, 2004). The issue of what is private space and what is public on the Internet has been particularly elusive. Questions about whether it is ethical to record a conversation in an online chat room and whether it is necessary to have a consent form for researching how participants interact in online public discussions have been particularly problematic (Hudson and Bruckman, 2005; Bos *et al*, 2009). Some researchers argue that these conversations are public but others say that participants own their own words. In social network sites such as Facebook the boundaries are particularly blurry. In some countries, many universities and national funding organizations take a conservative approach and insist that any work that involves human subjects should have ethical oversight. This is considered important because some users don't realize the potential dangers associated with putting private information online.

ACTIVITY 13.8

Studies of user behavior on the Internet may involve logging users' interactions and keeping a copy of their conversations with others. Should users be told that this is happening?

Comment

The answer can be controversial because, as mentioned above, some researchers believe that this information is public so it is not necessary to ask permission to analyze it. Others claim that informing people that their conversations are being analyzed changes the nature of the research. (But this is the case in most research as we discuss later.) We believe that it is better to tell users in advance that they are being logged. ■

DILEMMA

What would you do?

There is a famous and controversial story about a 1961–62 experiment by Yale social psychologist Stanley Milgram to investigate how people respond to orders given by people in authority. Much has been written about this experiment and details have been changed and embellished over the years, but the basic ethical issues it raises are still worth considering, even if the details of the actual study have been distorted.

The participants were ordinary residents of New Haven who were asked to administer increasingly high levels of electric shocks to participants when they made errors in the tasks they were given. As the electric shocks got more and more severe, so did the apparent pain of the participants receiving them, to the extent that some appeared to be on the verge of dying. Not surprisingly, those administering the shocks became increasingly disturbed by what they were being asked to do, but several continued, believing that they should do as their superiors told them. What they did not realize was that the so-called participants were, in fact, very convincing actors who were not being injured at all. Instead, the shock administrators were themselves the real subjects of the experiment. It was their responses to authority that were being studied in this deceptive experiment.

This story raises several important ethical issues. First, this experiment reveals how power relationships can be used to control others. Second and equally important, this experiment relied on deception. The experimenters were, in fact, the subjects and the fake participants colluded with the real scientists to deceive them. Without this deception the experiment would not have worked.

Is it acceptable to deceive subjects to this extent for the sake of scientific discovery? What do you think? ■

13.2.6 Evaluate, Analyze, Interpret, and Present the Data

Decisions have to be made about what data is needed to answer the study questions, how the data will be analyzed, and how the findings will be presented (see Chapter 8). To a great extent the method used determines the type of data collected, but there are still some choices. For example, should the data be treated statistically? Some general questions also need to be asked. Is the method reliable? Will the method measure what is intended, i.e. what is its validity? Will the evaluation study be ecologically valid or is the fundamental nature of the process being changed by studying it? Are biases creeping in that will distort the results? Will the results be generalizable, i.e. what is their scope? Each of these issues is explained below.

Reliability

The reliability or consistency of a method is how well it produces the same results on separate occasions under the same circumstances. Another evaluator or researcher who follows exactly the same procedure should get similar results. Different evaluation methods have different degrees of reliability. For example, a carefully controlled experiment will have high reliability, whereas observing users in their natural setting will be variable. An unstructured interview will have low reliability: it would be difficult if not impossible to repeat exactly the same discussion.

Validity

Validity is concerned with whether the evaluation method measures what it is intended to measure. This encompasses both the method itself and the way it is performed. If, for example, the goal of an evaluation study is to find out how users use a new product in their homes, then it is not appropriate to plan a laboratory experiment. An ethnographic study in users' homes would be more appropriate. If the goal is to find average performance times for completing a task, then a method that only recorded the number of user errors would be invalid.

Ecological Validity

Ecological validity is a particular kind of validity that concerns how the environment in which an evaluation is conducted influences or even distorts the results. For example, laboratory experiments are controlled so what the participants do and how they behave is quite different from what happens naturally in their workplace, at home, or in leisure environments. Laboratory experiments therefore have low ecological validity because the results are unlikely to represent what happens in the real world. In contrast, ethnographic studies do not impact the participants or the study location as much, so they have high ecological validity.

Ecological validity is also affected when participants are aware of being studied. This is sometimes called the Hawthorne effect after a series of experiments at the Western Electric Company's Hawthorne factory in the USA in the 1920s and 1930s. The studies investigated changes in length of working day, heating, lighting, and so on; however, eventually it was discovered that the workers were reacting positively to being given special treatment rather than just to the experimental conditions. Similar findings sometimes occur in medical trials. Patients given the placebo dose (a false dose in which no drug is administered) show improvement that is due to receiving extra attention that makes them feel good because they are getting attention.

Biases

Bias occurs when the results are distorted. For example, expert evaluators performing a heuristic evaluation may be more sensitive to certain kinds of design flaws than others, and this will be reflected in the results. Evaluators collecting observational data may consistently fail to notice certain types of behavior because they do not deem them important. Put another way, they may selectively gather data that they think is important. Interviewers may unconsciously influence responses from interviewees by their tone of voice, their facial expressions, or the way questions are phrased, so it is important to be sensitive to the possibility of biases.

Scope

The scope of an evaluation study refers to how much its findings can be generalized. For example, some modeling methods, like the keystroke level model and Fitts' Law (discussed in Chapter 15), have a narrow, precise scope. For instance, the keystroke level model predicts expert, error-free behavior so the results cannot be used to describe novices learning to use the system. (The problems of overstating results were discussed in Chapter 8).

Case Study 13.1

Evaluating YouTube

YouTube's mission is to create 'an online video community' that allows members to communicate about the videos on bulletin boards and via social-networking tools. A study was conducted by Rotman *et al* (2009) to evaluate a small part of YouTube. They chose to ask the question 'Is YouTube a community?' They answered this question using two different but complementary methods: a qualitative analysis using Grounded Theory (see Chapter 8) of users' feelings about the YouTube community and their interaction patterns, and a quantitative analysis of the actual YouTube structure as created by their articulated ties. They analyzed more than 30 YouTube videos and their findings showed how YouTube users were almost unanimous in their strong feelings about YouTube being an established community. They said that YouTube is a social space that offers the conditions needed to cultivate a community. Their own experiences revealed close-knit relations with others, for example:

Initially I just want to get my work seen. And then I started to get into communicating with people in this community I made really good friends, people I started talking to everyday, who I have known for most of a year. Some of which are very close friends I can talk to about personal things. (Participant 15)

I've made all kinds of friends, all through this site. Even people who live in this city that I would not have met otherwise. [I also] met people from across the world. (Participant 2)

Structurally, dense hubs of cross-relations typify communities whereas reports of solitary and interactions involving pairs are more characteristic of a broadcasting platform than a

community. In their study, users explicitly described the YouTube interaction pattern as that of a hub-like online community:

[the YouTube community is] *a spiderweb, because everyone interacts with each other, and everyone is a sender and a receiver. And this is me [points to himself] and I interact with all these different people.* (Participant 22)

A structural analysis was also conducted and showed the communication patterns among over 1500 users to see if a structure of community exists. The social network analysis revealed 2238 communication links between these users. Social networks exhibit a wide variation in their size and other structural characteristics. However, if a social network takes a particular form, which includes many clusters or hubs, we can assume that a community-structure exists. Alternatively, the network structure can be looser, indicating pair-wise or small group communication that does not support community interaction.

In structural network analysis several metrics can be used to describe the network, one of which calculates the amount of clustering in the network using a coefficient. A high clustering coefficient indicates strong clustering and hub-like structure. In this study the clustering coefficient was very low at 0.44, indicating that there does not appear to be a cohesive community structure and that the links are probably mostly between random pairs. The ties among YouTube users definitely did not exhibit a dense hub-like structure of a community.

In sum, the users reported feelings of membership, attachment to other users, fulfillment and influence through shared goals, and a feeling of belonging to the larger social body of YouTube contributors, referring to themselves as ‘YouTubers.’ However, the structural analysis demonstrated that users’ relationships do not reflect the perceived community; YouTube structure looks random. The discernable gap between the way users view their interaction patterns and the ways in which these interactions are manifested in the structural analysis indicates that different methods of analysis can provide different results. ■

ACTIVITY 13.9

The case study examined whether participants in an area of YouTube form a community. Two different methods were used in the study: an online ethnography that was analyzed using Grounded Theory, and a structural social network analysis. How adequately did it address: (1) Reliability, (2) Validity, (3) Ecological validity, (4) Bias, and (5) Scope.

Comment

1. The YouTube case study used two different methods: qualitative data collection and Grounded Theory analysis, and a structural analysis. The qualitative Grounded Theory

(Continued)

analysis has low reliability; different results might be produced by a different evaluator. The structural social network analysis is a quantitative method that would produce the same results if performed by another evaluator, i.e. it has high reliability.

2. Both methods have high validity; they are appropriate methods for the question that the evaluators wanted to address.
3. The researchers did not influence the behavior of the participants so the study has high ecological validity.
4. The qualitative Grounded Theory analysis is more likely to produce bias because the researchers might be influenced by their preconceived ideas when coding the data. The structural social network analysis is objective and is not prone to bias.
5. YouTube has an enormous repository of millions of videos and this case study examined a very small portion of it. Its scope is small. ■

Assignment

Find a journal or conference publication that describes an interesting evaluation study or select one from www.bcibib.org or from a digital library such as the ACM Digital Library. Then use the DECIDE framework and your knowledge from Chapters 7 and 8 to analyze it. Some questions that you should seek to answer include:

- (a) What are the goals and the questions that provide the focus for the evaluation?
- (b) Which evaluation methods are used?
- (c) What data is collected and how is it analyzed?
- (d) What practical and ethical issues have been considered?
- (e) Comment on the reliability, validity, ecological validity, biases, and scope of the study.
- (f) Is there evidence of one or more pilot studies?
- (g) What are the strengths and weaknesses of the study report? Write a 50–100 word critique that would help the author(s) improve their paper.

Summary

In this chapter we introduced the DECIDE framework, which will help you to plan an evaluation. There are six steps in the framework:

1. Determine the goals.
2. Explore the questions.
3. Choose the approach and methods.

4. Identify the practical issues.
5. Decide how to deal with the ethical issues.
6. Evaluate, analyze, interpret, and present the data.

Key points

- There are many issues to consider before conducting an evaluation study. These include the goals of the study, the methods to use, practical issues, ethical issues, and how the data will be collected, analyzed, and interpreted.
- The DECIDE framework provides a useful checklist for planning an evaluation study.

Further Reading

DENZIN, N. K. and LINCOLN, Y. S. (2005) *The Sage Handbook of Qualitative Research* (3rd edn). Sage Publications. This book is a collection of chapters by experts in qualitative research. It is an excellent resource.

HOLTZBLATT, K. (ed.) (2005) Designing for the mobile device: Experiences, challenges and methods, *Communications of the ACM* 48(7), 32–66. This collection of papers points out the challenges that evaluators face when studying mobile devices, particularly when the most appropriate study is a field study that may involve working in a different culture and changing physical environments regularly.

LAZAR, J., FENG, J. H. and HOCHHEISER, H. (2010a) *Research Methods in Human–Computer Interaction*. John Wiley & Sons Ltd, Chichester, UK. This book provides a useful overview of qualitative and quantitative methods. Chapter 14, ‘Working with Human Subjects,’ discusses ethical issues of working with human participants and so is particularly relevant to this chapter, but there is also much useful advice throughout the text.