Chapter 2
Qualitative Human Factors Research Methods

2.1 Human-Centered Design

As the name implies, human-centered design (HCD) is a framework for placing user goals, needs, capabilities, and limitations at the center of the product design process. Individual manufacturers will follow their own proprietary design processes, so HCD will be implemented differently in each case, but the following three tenets will be the same (Gould & Lewis, 1985):

1. Early and constant focus on the users and their tasks
2. Reliance on human–system performance data to guide design
3. Iteration

This HCD approach is distributed throughout the design process—not just during initial stages of need finding and requirements gathering but also through iterative refinement and post-market research and surveillance (see Fig. 2.1). The user, rather than the product, takes center stage in all parts of the process, requiring us to research the user’s job, tasks, workflow, needs, and preferences.

Focusing on the user throughout the whole design process requires a great deal of research. At the beginning you need to investigate and analyze the user’s needs, tasks, environments of use, and other factors. With each design iteration you need to inspect the user interface and conduct usability tests to identify usability problems. Even after the product is released, you should monitor it for use-error, and conduct competitive usability research, to learn how your product stands against the competition.

This chapter discusses the research, usability inspection, and data analysis methods typically employed during HCD. Not surprisingly, the list is long, and the trick for the researcher is to choose the right methods at the right time. As a result, these methods should be thought of as a toolkit rather than a recipe. Depending on the questions you need to answer, you will choose different research methods, and different data analysis techniques.
Human factors engineering (HFE) research involves gathering, analyzing, and interpreting data. Of course, people gather and interpret data every day, but we usually do so intuitively, and non-systematically. Research, on the other hand, answers questions in ways that are objective, systematic, and repeatable.

By “objective,” we mean that conclusions are based on data rather than intuition or opinion. Intuitions vary from person to person, and even moment to moment. To be sure, intuition does have a place in design, but intuition improves with increasing knowledge and information provided by research. It is risky to the point of foolishness to base multimillion-dollar design budgets solely on intuition.

Unlike intuition, which happens naturally and with little effort, objectivity is difficult, unnatural, and takes discipline. So, to facilitate objectivity we employ specific research methods designed to answer questions while reducing bias (Cacioppo & Freberg, 2013). We will discuss many of those methods in this chapter.

One note of caution is appropriate here. The primary purpose of research is to answer questions. We are often surprised by clients and students who begin research projects without specific research questions in mind. This can be an expensive use of time, money, frustration, and even professional reputation. Identifying, in detail, the questions you will address makes planning, conducting, and analyzing the study much easier. Undoubtedly, you will serendipitously answer other unexpected questions along the way, but identifying your questions ahead of time is critical. If you have no clearly defined research question, you are not doing research!

2.3 Reliability and Validity

Let us turn our attention to two important characteristics of research measurement—reliability and validity. Imagine you stood on your bathroom scale and looked at your weight. Now imagine doing it again, and then yet a third time. You would expect the weight to be just about the same, within a very small degree of variation. That is reliability—the consistency of an observation or a measure. If weighing
yourself resulted in wild swings of 15 or more pounds, you would no longer trust your scale.

Now imagine getting on the same scale with the knowledge that your usual weight is around 150 pounds. The scale, however, reads 275 pounds. As before, you reweigh yourself several times, and several times it returns the weight, 275 pounds. This scale is reliable; you always get the same result. But it is not valid. That is, it is not good at measuring what it is intended to measure: your actual weight.

As in scales, in HFE we need research observations and measurements that are both reliable and valid. On a questionnaire, say, under the same circumstances, we would expect the same person to express the same responses. This would suggest that our questionnaire is reliable.

A useful metaphor for the relationship between reliability and validity is an archery target. Think of the center of the target (the “bullseye”) as the concept or construct you are trying to measure. These concepts might be something like usability, interest in your product, task completion time, or error frequency. For each person you measure, you are taking a shot at the target. If you measure the concept perfectly for a person, you are hitting the center of the target. If you do not, you are missing the center. The more “off” you are for that person, the further you are from the center. Figure 2.2 shows three possible situations.

In the left-hand target, you are hitting the target consistently, but you are missing the center of the target. In other words, you are consistently and systematically measuring the wrong value for all respondents. This measure is reliable, but not valid (it is consistent but wrong). The middle scenario shows a case where your hits are spread across the target and you are consistently missing the center. Your measure, in this case, is neither reliable nor valid. The right-hand figure shows the Robin Hood or William Tell scenario; you consistently hit the center of the target. Your measure is both reliable and valid.
The fields of measurement and psychometrics have identified and described various types of reliability and validity, which are too detailed to cover here, but it is important to gather evidence that the information you are collecting is both reliable and valid. That is, can you count on it?

### 2.4 Selecting Research Participants

Since it is impossible to study every member of a population, we need to focus on only a representative subset of people, called a sample, which is used to make judgments about the entire population. In journal publications (and elsewhere), a study’s sample size is noted with an italicized \( n \). For example, \( n = 20 \) means the study’s sample consists of 20 participants.

The best way to ensure a representative sample is through random sampling, in which every member of the population has an equal chance of being selected for the study (Fig. 2.3). The more representative the sample, the better the results will generalize to the population. Unfortunately, random sampling is not always possible. Instead, HFE often uses samples of convenience, or groups of people who are easily accessible to the researcher. For example, many of the behavioral and medical sciences are criticized for using participants who are WEIRD (Henrich, Heine, & Norenzayan, 2010), that is, Western, Educated, and from Industrialized, Rich, and Democratic countries. As a result, it is possible that many findings in various academic journals would not apply to people with other characteristics.

### 2.5 Ethical Standards

It is critical to protect research participants from harm. Usually, HFE practitioners ensure this by adhering to guidelines set by the American Psychological Association (APA), which state that researchers carry out investigations with respect for the people who participate and with concern for their dignity and welfare (American Psychological Association, 2017). Often researchers use an institutional review board (IRB) to review the materials, procedure, and confidentiality measures provided by the study. At the core of ethical standards for human research is the idea that participation is voluntary. No participant should be coerced into participating. Researchers must obtain informed consent from all participants, and briefly describe the goals of the project, the potential risks and/or benefits, the procedures for

---

Fig. 2.3  Selecting research participants
maintaining confidentiality, and the incentives or payments offered. Once this information has been communicated, participants can agree to participate in the study.

The researcher should also make sure to do no irreversible harm to participants. Research using human participants should be private and confidential. Privacy refers to the participants’ control over how their information is shared, and methods for ensuring privacy are usually indicated in the informed consent. Confidentiality refers to the participants’ rights to not have their data revealed to others without their permission. Confidentiality is usually maintained by substituting codes for names and storing data in locked cabinets. APA guidelines are listed below.

1. Do no harm
2. Accurately describe risks to potential participants
3. Ensure that participation is voluntary
4. Minimize discomfort to participants
5. Maintain confidentiality
6. Do not unnecessarily invade privacy
7. Use deception only when absolutely necessary
8. Provide debriefing to all participants
9. Provide results and interpretations to participants
10. Treat participants with dignity and respect
11. Allow participants to withdraw at any time for any reason

2.6 Literature Review

Often, you can answer some of your research questions without collecting any new data at all by conducting a literature review. Many times, others have conducted research that is relevant to your current project, informing you about what is known, and what has been done. This information is usually available in journal articles, books, and conference proceedings, but it is also helpful to look at previous projects your organization has already completed. Hundreds of thousands of dollars can be invested in research investigations (e.g., contractor fees, opportunity cost, participant compensation, facility rentals). It is worth taking some time to check whether similar problems have been addressed through other research.

The literature review is intended to distill information from all sources, capturing the essence of previous research or projects as they might inform the current project. The review need not summarize everything from each source, but should synthesize the work by drawing connections among previous findings.

Internet resources such as search engines, digital journals, e-mail, and interlibrary loans expedite literature searches significantly, allowing researchers to access libraries from around the world. However, it is important to be discerning, ensuring that the literature selected for inclusion is not only relevant, but also from credible sources. Particular caution should be exercised if including website or blog resources, which may not be vetted or peer-reviewed for credibility.
2.7 Case Study

A case study is an in-depth analysis of a small number of events or people (Cacioppo & Freberg, 2013). Case studies are common in medicine, law, and business, and are sometimes used by human factors engineers when large numbers of participants are not available. For example, imagine the need to learn about people with a very rare disease (say an incidence of 1 in 200,000). It might be helpful to start by studying one individual at a time. One case will likely lead to hypotheses for future case studies or larger studies with more people. Because so few participants are involved (often only one), it affords the opportunity to gain detailed information about that one participant, even applying several research methods to more fully examine the person, small group, or event of interest.

2.8 Naturalistic Observation

Naturalistic observation is an in-depth study of potential users in their natural setting. A classic example of naturalistic observation in primatology was Jane Goodall’s long-term study of chimpanzees in Tanzania beginning in 1960. Though our observations may not take us to Tanzania, there is a wealth of insight to be gained from studying medical device users in their natural habitats, so to speak. Observing surgical techs in the surgical suite or physician’s assistants in the exam room provides insight into how they interact with the environment, including what materials they use, who they communicate with, what forms they fill out, what makes them nervous, which tasks are most frustrating, and so on. This is information not easily uncovered using some other research techniques (Fig. 2.4).

Even though naturalistic observation is unstructured, it is still critical to begin with research questions. These questions dictate the goals and approach for your work. In preparation, you should be able to answer the following questions before starting your observations.

- What questions do I intend to answer?
- Who (what work roles) would I like to observe?
- What activities would I like to observe?
- What communications would I like to observe?
- What devices or tools would I like to observe being used?
- What support materials (e.g., Instructions for Use) should I pay attention to?

Along the way, you will likely learn about:

- Responsibilities—for example, who does what and when
- Sequences of activities
- What problems can occur
- What activities are difficult and why
- How people troubleshoot problems
2.8 Naturalistic Observation

- How these problems are addressed
- How a task or activity is initiated
- How people know when they have completed the task or activity correctly

Observational research produces so much data that it is difficult to know what information to capture. Rick Robinson, Ilya Prokopoff, John Cain, and Julie Pokorny (Robinson, 2015) had the same challenge in their observational work. To organize research questions and information, they relied on a mnemonic using all English vowels: Activities, Environments, Interactions, Objects, and Users.

- (A) Activities: Describe sequences of actions that enable people to achieve a goal. It is worth documenting not only the activity itself but also the sequence that comprises that activity.
- (E) Environments: Describe the environment in which activities take place. Make sure to describe crowding, organization of the space, traffic flow, lighting, noise heating, and interruptions.
- (I) Interactions: Describe the interactions among people. Who talks to whom? What information do they exchange? When do people help others with their tasks? What types of questions are asked? What other media are used to interact? Do people discuss things only in person or also by telephone, e-mail, Slack, or Zoom? How do these interactions differ depending on the medium? Be sure to capture, where possible, both verbal and non-verbal communication.
- (O) Objects: Describe the tools, devices, forms, notes, and other artifacts that are used. How do these objects relate to the user’s goals, activities, interactions, and work in general?
24

2.9 Design Ethnography

Design ethnography adapts traditional ethnographic methods used in anthropology to study the social life of various cultures. It relies on observational research of people in their natural environments, but it also involves unstructured interviews and document analysis. Particular attention is paid to characterizing users’ lives, behaviors, language, and artifacts.

Design ethnography is different from traditional ethnography in that it is shorter and tends to be more focused. Traditional ethnographic studies in anthropology can last for years, whereas a typical design ethnography session will last only a few hours. As in all research methods, ethnography begins with a research question. As Fetterman (2019) pointed out, “The ethnographer enters the field with an open mind, not an empty head” (p. 1). Design ethnography is different from naturalistic observation because it relies on gathering all kinds of data. It uses not just observation, but interviews, artifacts (tools, etc.), photographs, sketches, document analysis, investigation of rituals, and so on. In other words, design ethnography is not limited to any one type of data collection method. What really matters is gaining insight, usually in the context of use.

2.10 Interviewing

Interviews fall into one of three approaches: structured, semi-structured, and unstructured. Each approach has different intentions, advantages, and disadvantages. The following sections describe each one, as well as a few tips about when it may or may not be appropriate to use them.

Structured Interview

A structured interview involves asking participants a specific set of questions, in a specific order. Every participant experiences the exact same procedure. As a result of this systematic approach, structured interviews often produce results intended for
quantitative analysis: the frequencies of times a participant “agreed” vs. “disagreed” on a topic, for example.

Structured interviews require considerable planning to identify the questions to be asked, and the order of asking them. Furthermore, researchers must use techniques to reduce participant fatigue—which can set in easily during longer structured interviews. These might be in the form of frequent breaks, or introducing other activities (e.g., surveys, questionnaires) to break up the monotony of a one-sided conversation.

**Semi-Structured Interview**

Semi-structured interviews are the most common approach taken in HFE. It represents the middle ground between a “structured interview” and an “unstructured” one, allowing researchers to follow a core set of questions, and fill in gaps as they are identified (Fig. 2.5).

The semi-structure gives the moderator freedom to go off-script to explore areas of interest based on the conversation, as long as this new direction remains in scope for the study objectives. This can be a challenging task for inexperienced study moderators. It is especially important that they define the scope of the interview beforehand.

Semi-structured interviews are useful for investigating the specific steps that users complete with a device. For example, imagine trying to understand a Sterile Processing Technician’s intake procedure for a used medical device. The “structured” part of the interview might come from trying to create a linear sequence of tasks. The “unstructured” part comes into play when clarification is needed.

One important part of designing a semi-structured interview is to build in enough time to permit the moderator to go off-script. Inexperienced researchers tend to design their protocols and moderator guides with too many pre-baked questions. They do not realize that answering these questions is time consuming.

**Unstructured Interviews**

Unstructured interviews have no predefined questions or topic order. Often, the moderator begins with a simple, open-ended question, and the interview (hopefully) progresses naturally from there. This is best done by a study moderator who is an expert—or, at least very knowledgeable—on the topic at hand. A moderator with expertise knows where gaps, uncertainties, or inconsistencies exist in product opinions or workflows. They intuitively lead the conversation in these directions to ferret out the information they need. Additionally, an expert can often spot another expert (or non-expert) based on the questions they ask. Thus, if the study moderator asks a
bunch of “softball” questions, the participant is less likely to explain important
details that they anticipate may be outside the moderator’s purview.

Yet even when the moderator is an expert in the field, unstructured interviews are
rare. There are simply too many opportunities for the interview to go astray, includ-
ing situational factors such as the participant being tired after a long day of work.
Stanton, Salmon, and Rafferty (2013) also point out that unstructured interviews
have the potential for crucial information to be neglected or ignored.

**Interview Questions**

There are two basic types of interview questions: closed-ended and open-ended. A
closed-ended question presents the participant(s) with a limited set of response
options. This might be something as simple as a “yes” or “no” to a question such as:
“Do you feel that other healthcare professionals would be able to use this product
safely?” Additional response options can also be incorporated, such as survey ques-
tions to assess a participant’s agreement or disagreement with a statement.

Closed-ended questions are underappreciated, probably because they create bot-
tlenecks in a line of questioning. That is, the participant must first choose a “yes” or
“no” response, then the interviewer can ask a follow-up question they really care
about—why or how? In some cases, having a participant commit to a choice pre-
pares them to justify their selection. However, it may be better to save time and just
get to the question you intended to ask in the first place.
You can often tell that a question is closed-ended if it begins with one of the following:

- Can you…?
- Do you…?
- Will you…?
- Have you…?

These can always result in a Yes or No response. Of course, in natural conversation, most people know to elaborate more. For example, a question such as, “Can you show me how you did that again?” will almost always be returned by the participant with an action, rather than the simple response of, “yes, I can.”

Closed-ended questions must be thought through carefully before use. They are so common in everyday speech that we often do not realize we are using them with participants in the first place. And, try as we might, a closed-ended question will slip its way into your interview whether you like it or not. This is to be expected given the interviewer’s responsibility to lead the interview, remember the important details, take notes, and stay engaged with the participant. To limit these slip-ups, the interviewer should offload as many tasks to note takers and fellow observers as possible. They should also complete several “dry runs” with pilot participants and colleagues to identify areas where they may get stuck.

An open-ended question enables participants to explain things in their own terms, and to elaborate. That is, they are not bound by the choices or options presented by the study moderator or testing materials. Open-ended questions often begin with one of the following:

- How…?
- Why…?
- What…?

Keep in mind that open-ended questions do not actually have to be questions at all. In fact, just telling the participant what you want them to talk about is perfectly acceptable, and, in many cases, more natural during conversation. For example, a good go-to statement to prompt an open-ended response is one of the following:

- Tell me (more) about….
- Walk me through….
- Explain how….

There are several benefits of using open-ended questions. As Stanton et al. (2013) point out, they allow the participant to answer the question in whichever way they choose, often resulting in more pertinent information than you would receive through a closed-ended question. And, perhaps just as important, open-ended questions keep the conversation moving forward. By contrast, a closed-ended question really only has one follow-up question: “Why did you pick that choice?” Open-ended questions are also more conversational and can feel less intimidating. It feels like there is less of a spotlight on them, or less like they are being “tested.”
Despite these benefits, one drawback to open-ended questions is that they can take longer to analyze. You cannot simply count up who chose what. The researchers may have to incorporate qualitative analysis techniques, such as affinity diagramming, to tease apart similar and dissimilar concepts.

**Ensuring Interviews Are Productive**

Below are a few tips for productive interviews.

- **Make participants comfortable:** Interviews are unfamiliar and frankly an odd situation for most participants. The participants will vary in their degree of extroversion, among other traits, so some may find it intimidating. Your ability to put them at ease will improve their responses and provide you with better information.

- **Start broad and funnel to more specific questions:** It is often helpful to think about the structure of an interview as a funnel with general questions at the top and asked at the beginning, then funneling down to more specific questions as the interview proceeds. This enables the participants to express what is top of mind first, then follow with more specific issues.

- **Avoid leading questions:** Leading questions encourage the answer you are interested in. We once observed a software developer ask a participant after a usability test, “that was not too unusable, was it?” That is an example of a leading question.

- **Ask follow-up questions as the interview unfolds:** The first answer from the participant is not usually the most insightful one. We often use laddering (Reynolds, Dethloff, & Westberg, 2001), which requires us to ask “Why” questions in succession after the first response. This enables us to understand the reasoning behind the first, superficial, answer.

- **Do not be afraid of silence:** If you need a moment to gather your thoughts, it is perfectly acceptable to tell a participant that you need a second to think about how to phrase your next question. Silence is also a natural cue that the person speaking last (e.g., the participant) should elaborate more. Participants themselves might need a moment to process their thoughts before speaking.

### 2.11 Focus Groups

A focus group is an interview with several participants at the same time, either in person or remotely. This method lends itself to covering a wide range of topics, ranging from the perceived usability of a system, common workflows followed by users, or to simply elicit feedback on a new design concept (Stanton et al., 2013). One benefit of the focus group is that it enables group interactions, allowing
participant responses to build upon each other (Farnsworth & Boon, 2010). It also allows large amounts of qualitative data to be collected quickly.

Traditionally, focus group samples are composed of homogeneous groups of participants (i.e., all from the same background). This might include subject matter experts (SMEs), C-suite executives, or people with Type II diabetes who use continuous glucose monitors (CGMs). Topics usually follow a semi-structured interview format to allow time for open discussion.

Less commonly, focus groups are made up of heterogeneous groups (i.e., people from different backgrounds). Heterogeneous groups can be useful in the early requirements-gathering phases of designing a medical device, as specific end-user groups may not be defined yet. This approach can also be useful when multiple user groups work with the same medical device, but may not work with each other on a regular basis. For example, a respiratory therapist and a reprocessing technician could both work with the same flexible endoscope in an average day, but may not have direct contact with each other. As such, they may not understand the constraints and issues associated with each other’s experience with the device. Sometimes having one user group hear the “pains” of another can generate opportunities for redesign.

Although this approach generates a lot of data quickly and easily, it is prone to a variety of social biases (e.g., groupthink behaviors) and may not be valid or reliable. Technologies such as audience response systems—also known as clickers to university students—or similar smartphone apps can be a work-around to this issue. These platforms allow focus group participants to cast their selections privately, thereby reducing the likelihood of bias. In general, however, focus groups have progressed toward the more qualitative end of the spectrum—especially in the field of medical HFE.

Klein, Tellefsen, and Herskovitz (2007) point out that a focus group can be completed in one of four ways, though only three of these combinations actually make sense from a logistical standpoint. These variations are outlined in Table 2.1.

**In-Person Focus Groups (Synchronous, Co-Located)**

An in-person focus group is the traditional format that people imagine when they hear the term focus group. Usually, it involves a large room with a conference table surrounded by seats so participants can see each other while conversing (Fig. 2.6).

| Table 2.1 | Variations of focus group formats |
|-----------|----------------------------------|
| Synchronous, co-located | Asynchronous, co-located (not common) |
| At the same time, in the same location | At different times, in the same location |
| Synchronous, distributed | Asynchronous, distributed |
| At the same time, in different locations | At different times, in different locations |
An in-person focus group should be led by at least one study moderator who is both trained and experienced. It is recommended that additional moderators should be considered for groups with more than eight people. The moderator keeps participants on track, and guides the progress at appropriate times. Not surprisingly, the study moderator has to be familiar with and kept apprised of the study objectives and research questions well in advance in order to do their job well and to elicit the most useful information out of the focus group.

In-person focus group sessions should ideally accommodate five to eight participants. The session length should be about 15–20 min per participant involved. For example, this would warrant that a five-participant focus group be approximately 90 min long (Stanton et al., 2013). The obvious caveat to this recommendation is that the topics of the study itself may demand more (or less) time. If the study requires participants to view a lengthy video, for example, the session length will likely change (Table 2.2).

**Remote (Online) Focus Groups (Synchronous, Distributed)**

Recently, focus groups have begun to make use of remote video conferencing platforms instead of conference rooms and lab environments. Unlike in-person focus groups, each participant in a remote focus group meets at a designated time online. This is especially useful for hard-to-reach participants (e.g., patients with rare disorders or specialized surgeons), as all they need to participate is a laptop, camera, and stable Internet connection. Additionally, remote focus groups can cost less (Reid & Reid, 2005; Summanen, Liikkanen, Laakso, & Leisti-Szymczak, 2013).
Remote focus groups can be slightly more challenging to moderate, since participants do not have access to the same non-verbal cues, such as facial expressions, hand movements, and verbal pauses that people naturally pick up on in-person. Furthermore, it can be challenging to ensure that all participants engage equally in the conversation. Some participants may be afraid to interrupt others or may choose to remain idle in the background until called upon. For these reasons, remote focus groups should be conducted with fewer participants than in-person groups. Despite these potential shortcomings, however, Abrams, Wang, Song, and Galindo-Gonzalez (2015) found online focus groups using webcams produced similar results (i.e., data richness, word count) to in-person focus groups following the same discussion topics (Table 2.3).

**Bulletin Board (Online) Focus Groups (Asynchronous, Distributed)**

A third form of focus group—the Bulletin Board Focus Group (BBFG)—also uses the Internet but forgoes audio and video components altogether. Instead, a BBFG (Fig. 2.7) is completed asynchronously over multiple days on a platform similar to an Internet forum-style website. Unlike a typical website forum, however, a BBFG is by invitation only, and is closed to public viewing and contributions.
A BBFG study begins by the moderator laying the groundwork and expectations for the study. This often involves a brief tutorial of the BBFG platform (i.e., how does it work?). Then, the moderator presents the first of several topics to participants, who can respond in writing, brief videos, or images. Depending on the platform settings, participant identities can be replaced with a pseudonym to promote anonymity. Likewise, the research team may opt to prevent participants from seeing others’ comments until after they have responded to the moderator’s initial inquiry. This can reduce social loafing and other biases throughout the study. It can also help create “lower stakes,” or judgment-free environment to assuage participant concerns about criticism from peers.
One benefit of a BBFG over in-person or remote focus groups is that several dozen people can participate in a single study. This can be a quick way to capture lots of opinions and information in one study effort. Bear in mind, however, that more information does not necessarily mean your study elicited higher-quality information. Similar to in-person and remote focus groups, having a well-trained study moderator at the helm can make the difference between low and high participant engagement (Table 2.4).

### 2.12 Diary Studies

A diary study asks research participants to record their daily experiences, activities, and events in writing or video. Diaries are useful for collecting information over a period of time when you cannot be there to observe directly. Nowadays, most diary studies are conducted online. Participants may be asked to document each time they engage in a particular behavior, encounter a product or situation, or have specific...

---

**Table 2.4 Advantages and disadvantages of Bulletin Board Focus Groups**

| “Pros” of Bulletin Board Focus Groups | “Cons” of Bulletin Board Focus Groups |
|--------------------------------------|--------------------------------------|
| • Elicits responses from several end-users in a single study session | • No audio/video of participants (unless study tasks are designed with that intent ahead of time) |
| • Sharing opinions and experiences can “connect the dots” in ways that individual experiences cannot provide | • The “richness” of hearing and seeing someone speak is lost |
| • Useful way to elicit information when team’s current knowledge on the subject may not be sufficient to create a survey or questionnaire | • You may get a “dominating” voice in the group that stifles collaboration or biases others |
| • Avoids wasting time in one-on-one interviews talking about the same issues with individual participants | • Quality of session depends heavily on the capabilities of the study moderator |
| • Participants do not have to think on the fly; they can think about their responses before replying to the group | • Data gathered do not lend itself well to common forms of statistical analysis |
| • Easier for hard-to-reach participants to commit to the study | • Some data collection methods are suspect to bias |
| • Participants can pick up and stop during the day at their leisure | • Difficult to redirect group when they get off-topic |
| • Ability to run more participants than other types of focus groups | • Inability to shift topics quickly on the fly |
| • (Usually) costs less than in-person focus groups | • Difficult to follow-up or clarify on specific topics or comments |
| • Anonymity helps reduce participant concerns over criticism or scrutiny | |
types of interactions. Other studies may require regular entries at particular times of day, or a log of items in summary at day’s end (Martin & Hanington, 2012) (Fig. 2.8).

A research method similar to a diary study—referred to as experience sampling (Hektner, Schmidt, & Csikszentmihalyi, 2006)—involves collecting data about behaviors, thoughts, or feelings from people at various times of the day. At random or fixed time intervals, participants receive texts querying them about what they are doing, how they are feeling, what is on their mind, and so on. Observations are often entered into an application resembling a diary or journal, and entries can include text, images, or recordings. This is different from a diary study, in which the participant is free to choose when they would like to submit their diary entry, which often makes diary studies more enjoyable and easier for participants to fit into their daily routines. However, it also comes with the risk of participants (occasionally) forgetting to complete their assignments.

Keep in mind that since these types of studies often run over several days, weeks, or months at a time, you may lose participants along the way—a concept referred to as attrition. While attrition is a common occurrence in any type of longitudinal study, in our experience you will lose more participants when there is little “skin in the game” for them to stay invested beyond monetary compensation. Remote diary studies lack the face-to-face contact, which reminds participants that there are actual people invested in this study with them. For these reasons, it is usually a good practice to over-recruit your diary study sample by 10–20% participants to help offset attrition. Otherwise, you run the risk of not meeting your study sample size objectives.
2.13 Critical Incident Technique

Originally used in aviation safety, the Critical Incident Technique (CIT; Flanagan, 1954) is really a set of research techniques. A commonly used approach is to ask seasoned users to describe a time period or event of significance—either good or bad. For example, you might ask the participant to tell you about a situation in which things almost went terribly wrong. What happened? Why did it happen? What did you do? How did you feel? How did you solve the problem? What do you think could have prevented it? Seasoned users’ responses to these questions are particularly helpful in the design of medical devices.

Similarly, you could ask about positive critical incidents. For example, “tell me about a time that you were really happy when using the device.” What happened? Why did it happen? How did you feel? and so on. Collecting a large volume of these stories (50–100) enables you to analyze each one according to cause and outcome, as well as the user’s responses, emotions, and motivations.

2.14 Participatory Design

Participatory design involves designers and users working together to develop solutions, ensuring that the product is useful, usable, and desirable. Participatory design appears to have started in the 1970s in Norway where computer scientists worked with ironworker and metalworker unions to integrate technology (Kuhn & Winograd, 1996). Participatory design has since expanded in scope and methods, gaining in design research. It encompasses several methods all dedicated to active consultation with users in the design process, ideally through face-to-face contact in activity-based co-design engagements.

There are a variety of participatory design methods used to elicit different types of information from research participants. For example in a recent study we had physicians select emotionally evocative images from a large database to describe their current and ideal healthcare experiences. In these exercises, the images serve as a vocabulary for participants to express how they feel. The images chosen are important, but we find the real value can be found in the reason participants provide for choosing the images they did. It often leads to rather rich discussions.

In Fig. 2.9, a participant sketches his ideal user interface for a medical device. Sketching involves using simple tools like large sheets of paper, markers, and Post-it Notes for participants to describe their ideal designs on paper. Again, the value of this exercise is often obtained from the interview discussing the artifact after it has been created.

Figure 2.10 shows the result of a Velcro modeling session. During these participatory design sessions, we provide a variety of ambiguous shapes (usually foam core), and objects covered in felt and Velcro. Participants are instructed to create a device that would help them achieve their ideal experience. As you can see, the
details can seem almost comical, but the reasoning behind them are informative. Follow-up interviewing enables us to probe:

- Which features were included, and which were not.
- The reasoning behind the placement of items.
- The relationships and grouping among items.
• The size of each item. For example, often participants will make the most important or most frequently used items larger than others.

2.15 Contextual Inquiry

Contextual inquiry combines observational and interview methods with a set of analysis methods to learn about users, their tasks, environments, and challenges. Its strength derives from its combination of observational and verbal response methods, its focus on context of use to understand how people work, and techniques for analyzing and visualizing the resulting data. As with observational methods, one goal of contextual inquiry is to observe users in their natural environments.

Contextual inquiry results can be used to identify needs, define requirements, improve processes, and learn what is important to users. These sessions can enable the researcher to document:

• Sequence of tasks
• Artifacts and tools people use to accomplish work
• Communication flows
• The impact and influence of the physical environment on the work
• Impact and influence of the culture on the work

Contextual inquiry achieves its goals by adhering to four principles:

• Focus: This refers to identifying the purpose of the research, the questions you are attempting to answer, and the information you seek to collect. This focus may change over time, but you always want to remember what your focus is at any one point. Any time you are surprised, or detect a contradiction, provides an opportunity to refocus inquiry.
• Context: This involves observing and interacting with users in their natural environment as they conduct their day-to-day activities. For surgeons, this environment will be the surgical suite, for home users this is likely to be the home. The fact that observers see the task as it occurs naturally means they can see the actual experience rather just hear about the user’s memory (a summary) of the experience.
• Partnership: In our opinion, partnership is one of the greatest strengths of contextual inquiry. This approach often employs a master/apprentice model, in which the user behaves as a master teacher and the researcher serves as an apprentice. This encourages the user to demonstrate what they are doing, think out loud, use accessible language, and expect naive questions from the researcher. It puts both of them on the same team.
• Interpretation: After data are collected, researchers spend significant time and effort to interpret the information collected, and potentially to verify their interpretations with the participants.
Beyer and Holtzblatt (1997) use two mnemonics—FACES and BASIC—to remember the types of information to attend to during contextual inquiry interviews.

- (F) Flow—refers to work transactions.
- (A) Artifacts—these are objects people use or produce during their work. Analyzing the artifact, including its location, whether it is public or private, whether it is well worn or pristine, and so on, at a later time can help provide insights into the user, goals, tasks, challenges, and culture.
- (C) Context—this can refer to pressures or constraints on the user.
- (E) Environment—physical layout of work space and work structures. This can reveal priorities, task sequences, and challenges related to things like storage, communication, and so on.
- (S) Sequence—this provides information about tasks, steps, task triggers (what tells us the task is ready to be conducted), the order in which they are conducted, goals task, and flow. This can help point out activities that might need support.
- (B) Breakdowns—situations in which things go wrong.
- (A) Anecdotes—stories about situations that have happened before.
- (S) Scenarios—situations that occur during observation.
- (I) Insights—information gained.
- (C) Custom tools—these are artifacts that users create themselves to improve the efficiency or effectiveness of their work.

2.16 Analyzing Qualitative Data

Overview

Obviously, user research can generate a great deal of data quickly. This raises the question of how to make sense of all this information. Beyer and Holtzblatt (1997) suggest that the research and design teams gather within 48 h of the contextual inquiry session for a 2-h interpretation session. During this session, the team members use meeting notes, artifacts, and other data to replay their visit, and transfer the information onto notes and work models. The goal is to share the visit information and to develop a shared understanding of the data. They start by assigning the following roles:

- Storyteller—who reviews the notes out loud and represents what they learned from the user.
- Note taker.
- Facilitator—who keeps the process on task. This person uses their team management skills to get data down and ensure that team members feel heard.
- Modeler—who captures and generates data models that illustrate what was learned.
During the interpretation sessions, it is important to avoid abstracting or summarizing the data until later. It is also important to avoid critiquing or arguing over the data at this point. Instead, teams should default to the interviewer for the final interpretation or mark it as a data question and follow-up with users later.

After initial interpretation sessions, it is time to consolidate the data into an Affinity Model. Affinity diagramming identifies patterns or clusters of data collected during the contextual inquiry. Researchers start by writing each observation, insight, quote, or idea on Post-it Notes (one observation per Post-it). Notes might include the following types, and be assigned the labels in Table 2.5.

Each Post-it Note is based on something the researcher directly observed or heard from participants. This might be an important quote, a part of the interface that gave someone trouble, or a couple words that address what happened (e.g., could not find the power button). Post-its are initially placed randomly on a wall or table (i.e., no affinity to one another). As more interviews or sessions are conducted over time, researchers will add more Post-its to their collection. After an agreed upon point, the researchers take stock of the assets they have created and begin to look for relationships or patterns among them.

Stronger and more reliable insights from affinity diagramming are generated by drawing information from multiple studies or observations, rather than just one. For example, Martin and Hanington (2012) recommend that affinity diagrams created from ethnographic studies should include interviews or observations from at least four to six different locations. Each interview or observation will usually yield a large amount (e.g., 50–100) of observations, depending on the length of the interview and the content addressed in it. As a result, a full round of interviews and observations will often provide researchers with several hundred items for their affinity diagram.

Effective affinity diagram content can be challenging to develop for inexperienced researchers. Often, novices will put too much information on a Post-it, or will

| Note label | Meaning                          |
|------------|----------------------------------|
| #          | Unique note number               |
| U#         | Unique user number               |
| What       | What is happening                |
| Why        | Why something happened. This depicts intent and importance |
| P          | User profile note                |
| WN         | Work note                        |
| BD         | Work breakdown                   |
| DI         | Design idea                      |
| I          | Major insight                    |
| H          | Data hole                        |
| C          | Context note                     |
| A          | Artifact note                    |
use language that is non-descriptive and hard to understand after the study is finished. Experienced researchers also know to incorporate additional factors into each Post-it, such as a participant identity (ID) number, the user group, and the study from which the information originated. These details can be vital after the fact to trace back the original data.

Keep in mind that affinity diagramming—like other types of analytic exercises and researcher-generated tools—is subject to bias. One way to reduce bias is to avoid sorting important criteria by Post-it color. This seems counterintuitive; after all, why not sort all the “nurses” in one color, and all the “physicians” in another? While group-color distinctions make it easier to see group patterns from afar, they can also lead researchers to read into these differences too much. It is easy to detect group differences when there really are none. This also encourages teams to think of problems as user group-specific, rather than determining whether the problem might apply to all—or, at least other—user groups (Fig. 2.11).

**Task Analysis**

Task analysis (Kirwan & Ainsworth, 1992) studies what a user (or team) must do (actions and cognitive processes) to achieve their goals using a system. It provides a step-by-step description of each important task conducted with a device. Its purpose is to understand and represent human and device activities, human–device interactions, and performance in a specific task or scenario. There have been over

*Fig. 2.11* Affinity diagramming using photos as reference (This exercise was conducted in March 2020 during the COVID-19 pandemic)
100 task analysis techniques described in the literature (Diaper & Stanton, 2004), which makes it a bit hard to determine which one is best for your project. Because of this variety, we will just cover one, probably the most common, here—hierarchical task analysis (HTA; Stanton, 2006).

HTA analyzes the goals and tasks of a system, treating those goals and tasks as a hierarchy. Specifically, every task is a means of achieving some goal. HTA starts by describing the purpose of the system at the highest level. It then lists the highest-level tasks (and their concomitant goals) required to achieve that purpose. Once it has done that, it further decomposes those tasks into subtasks and subgoals. This means that in order to satisfy the goal in the hierarchy its immediate subgoals have to be satisfied, and so on (Stanton, 2006).

Annett (2004) described the process as continual reiteration and refinement. First, you describe a rough outline of the hierarchy. Then you decompose those tasks and goals into a subgoal hierarchy. Then you do the same thing again for that hierarchy, and so on. Theoretically, this could go on forever; in fact, knowing when to stop the analysis is “one of the most difficult features of task analysis” (Annett, 2004). Eventually, you reach a subtask level that is granular enough to analyze. At this point, breaking down the subtask any further begins to feel almost ridiculous. We wish there were a more scientific way to describe this, but that is really how it works. In fact, Annet (2004) urges task analysts not to pursue re-description unless it is absolutely necessary. That is, once you reach a satisfying and useful description, you will reach a point of diminishing returns. Stop there (Fig. 2.12).

The flowchart resulting from HTA provides an exhaustive description of the activity, which can serve as a precursor to other HFE analysis methods. Each terminal subtask (the leaves of the tree) in that flowchart can then be analyzed according to:

- Criticality of each subtask
- Ease or difficulty of completing the subtask
- Stimuli that initiate step
- Actions required

![Flowchart resulting from a hierarchical task analysis](Fig. 2.12)
• Criteria for successful performance
• Interrelationships between people, objects, and actions
• Decisions that need to be made
• Information needed to make those decisions
• Feedback requirements
• Potential problems
• Potential use-errors
• Potential interruptions
• What mistakes might the user make?
• What must the user do when they make a mistake?
• What strategies do the users employ to improve their performance?
• Whether the next subtask is immediately obvious?
• Whether the series of steps is efficient?
• What must the user remember between subtasks?

The advantages of task analysis are numerous:

• It enables you to create a structure that helps make sense of all the observation and interview data
• It makes the data useful for design or training
• It helps identify the goals and functions of the system
• It highlights critical steps in the users’ tasks
• It identifies problems with the ways that users conduct their tasks
• It helps us to learn more about the user’s environment
• It shows how a new system might fit in with old ones
• It shows how a new system will fit with the workflow
• It inspires new design ideas

The disadvantage is that it is time consuming, and can be a bit unreliable. Because it is a qualitative analysis method, different analysts may develop different flowcharts and analyses. Anecdotally, researchers often report that the value of task analysis resides in actually conducting the activity rather than the flowchart and analysis table that results. You learn a great deal about the system (perhaps more than anyone else in the company) when you conduct the task analysis. This is important if you are involved in design.

**Swimlanes**

Swimlane diagrams (Fig. 2.13) illustrate the activities of multiple users in a flow diagram, providing a bird’s-eye view of who does what and when. The top lane is reserved as a storyboard lane to illustrate the events in an activity or process. A user experience lane (or set of lanes) uses a flowchart of boxes and arrows to depict the users’ activities. A business process lane describes how business goals and processes...
fit in with the activities, and finally a tools and systems lane documents the other devices and systems materials that would be used along the way.

**Journey Maps**

Journey maps (Fig. 2.14) illustrate the experiences people have when interacting with a product or service. It includes the actions, decisions, feelings, perceptions, and frame of mind—including the positive, negative, and neutral moments (Martin & Hanington, 2012).

The map should represent an experience, including moments of indecision, confusion, frustration, as well as delight. Multiple maps will need to be created for multiple roles, as each role will have different tasks and goals, and will experience different breakdowns and successes on their journey.
Scenarios

You can use some of the qualitative data you have collected to form scenarios. A scenario (Carroll, 2000) is a story describing the future use of a product from the user’s point of view. This is done to make design ideas explicit and envision the ways the device will be used. It is also helpful to depict exactly what happens in suboptimal situations, such as high stress, high noise, and when time constrained. Scenarios should follow a traditional story arc, beginning with a trigger, preconditions, discussing activities, and providing resolution.

User Profile

Often the qualitative data you collect will tell you about various types of users who use your device. A user profile lists the most important characteristics of your users, so that they can be accommodated in the device design. For example, consider the user profile of an electrophysiologist using a programmer device for a pacemaker/defibrillator. That profile might describe the user as a well-educated medical professional, who is a good reader, medically literate, adept at reading graphs, and between
the ages of 25 and 60. The profile might go further to indicate that they are time
constrained, dislike reading instructions for use, are highly motivated to learn, and
technologically savvy. Note that the same device may have several user groups, and
thus would require several user profiles. For example, a technician who interrogates
the device may have different characteristics from the electrophysiologist who ulti-
mately uses the data.

Many variables contribute to how users or different user groups perform with a
device, and it is up to you as the researcher or designer to determine which variables
are most important for performance and satisfaction with your device. Some sample
variables are listed below.

- Job type
- Reading ability
- Attitude toward technology
- Motivation to learn
- Typing skill
- Height
- Task experience
- Training method
- Application experience
- Native language
- Computer literacy
- Frequency of use
- Turnover rate
- Education level
- Major in college
- Knowledge and experience
- Tasks and needs
- Mandatory or discretionary use
- Task or need importance
- Task structure
- Social interactions
- Primary training
- Attitude and motivation
- Patience
- Stress level
- Expectations
- Vision, hearing, cognitive processing, manual dexterity
- Handedness
- Disabilities
Prototyping

A prototype is a model of a product that enables you to demonstrate what the product looks like, how it is organized, or how it works. Early prototypes usually have low fidelity or low resolution, just demonstrating key aspects of the design. These early prototypes usually do not work; that is, they do not actually perform a medical function. On the other hand, they might appear to work. For example, a prototype of a software user interface might switch from one screen to another when a particular button is pressed. This may enable a research participant to interact with the prototype like it was the real product. This can be helpful during feedback sessions or usability testing. Over time, and with additional development, the design team may develop a high-fidelity prototype, which closely resembles the end product and may even include some basic functionality.

Resources

• Holtzblatt, K., & Beyer, H. (2016). Contextual design. In Design for life (2nd ed.). San Rafael, CA: Morgan & Claypool.
• Laurel, B. (2003). Design research: Methods and perspectives. Cambridge: MIT Press.
• Martin, B., & Hanington, B. (2012). Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Beverly, MA: Rockport Publishers.

References

Abrams, K. M., Wang, Z., Song, Y. J., & Galindo-Gonzalez, S. (2015). Data richness trade-offs between face-to-face, online audiovisual, and online text-only focus groups. Social Science Computer Review, 33(1), 80–96.
American Psychological Association. (2017). Ethical principles of psychologists and code of conduct (2002, amended effective June 1, 2010, and January 1, 2017). Retrieved August 24, 2020, from http://www.apa.org/ethics/code.
Annett, J. (2004). Hierarchical task analysis (HTA). In Handbook of human factors and ergonomics methods (pp. 355–363). Boca Raton, FL: CRC Press.
Beyer, H., & Holtzblatt, K. (1997). Contextual design: Defining customer-centered systems. Amsterdam: Elsevier.
Cacioppo, J. T., & Freberg, L. A. (2013). Discovering psychology: The science of mind. Boston, MA: Wadsworth Cengage Learning.
Carroll, J. M. (2000). Making use: Scenario-based design of human-computer interactions. Cambridge: MIT Press.
Farnsworth, J., & Boon, B. (2010). Analysing group dynamics within the focus group. Qualitative Research, 10(5), 605–624.
References

Fetterman, D. M. (2019). *Ethnography: Step-by-step* (Vol. 17). Washington, DC: SAGE Publications, Incorporated.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327.

Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM, 28*(3), 300–311.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (Eds.). (2006). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage Publications.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466*(7302), 29–29.

Kirwan, B., & Ainsworth, L. K. (Eds.). (1992). *A guide to task analysis: The task analysis working group*. Boca Raton, FL: CRC press.

Klein, E. E., Tellefsen, T., & Herskovitz, P. J. (2007). The use of group support systems in focus groups: Information technology meets qualitative research. *Computers in Human Behavior, 23*(5), 2113–2132.

Kuhn, S., & Winograd, T. (1996). Participatory Design. In T. Winograd (Ed.), *Bringing design to software*. Boston: Addison-Wesley.

Martin, B., & Hanington, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Beverly, MA: Rockport Publishers.

Reid, D. J., & Reid, F. J. (2005). Online focus groups: An in-depth comparison of computer-mediated and conventional focus group discussions. *International Journal of Market Research, 47*(2), 131–162.

Reynolds, T. J., Dethloff, C., & Westberg, S. J. (2001). Advancements in laddering. In T. J. Reynolds & J. C. Olson (Eds.), *Understanding consumer decision making: The means-end approach to marketing and advertising strategy* (pp. 108–134). UK: Psychology Press.

Robinson, R. E. (2015). Building a useful research tool: An origin story of AEIOU. In *EPIC: Advancing the value of ethnography in industry*. Retrieved July 29, 2020, from https://www.epicpeople.org/building-a-useful-research-tool.

Stanton, N., Salmon, P. M., & Rafferty, L. A. (2013). *Human factors methods: A practical guide for engineering and design*. UK: Ashgate Publishing, Ltd.

Stanton, N. A. (2006). Hierarchical task analysis: Developments, applications, and extensions. *Applied Ergonomics, 37*(1), 55–79.

Summanen, I., Liikkanen, L. A., Laakso, M., & Leisti-Szymczak, A. (2013). Online focus groups: A new tool for concept testing with remote users. From the Proc. of Co-Create 2013. Espoo, Finland. 16–19 June (pp. 685–696).