

Introduction

1

The goal of this book is to show how usability metrics can be a powerful tool for successfully evaluating the user experience for any product. When some people think about usability metrics, they feel overwhelmed by complicated formulas, contradictory research, and advanced statistical methods. We hope to demystify much of the research and focus on the practical application of usability metrics. We'll walk you through a step-by-step approach to collecting, analyzing, and presenting usability metrics. We'll help you choose the right metrics for each situation or application, and show you how to use them to produce reliable, actionable results without breaking your budget. We'll give you guidelines and tips for analyzing a wide range of usability metrics and provide many different examples of how to present usability metrics to others in simple and effective ways.

Our intention is to make this book a practical, how-to guide about measuring the usability of any product. We aren't going to give you a lot of formulas; in fact, there are very few. The statistics will be fairly limited, and the calculations can be done easily in Excel or some other common software package or web application. Our goal is to give you the tools you need to evaluate the usability of any product, without overwhelming you with unnecessary details.

This book is both product- and technology-neutral. The usability metrics we describe can be used for practically any type of product and any type of technology. This is one of the great features of usability metrics: They aren't just for websites or any single technology. For example, task success and satisfaction are equally valid whether you evaluate a website, a treadmill, or a toaster. More advanced technologies, such as websites, mobile phones, software, and consumer electronics, are of special concern because they're generally more complicated, but the basic premise remains the same.

The "half-life" of usability metrics is much greater than any specific design or technology. Despite all the changes in technology, the metrics essentially stay the same. Some metrics may change with the development of new technologies to measure usability, but the underlying phenomena being measured don't change. Eye-tracking is a great example. Many researchers wanted a method for determining

where exactly a user is looking on the screen. Now, with the latest advances in eye-tracking technology, measurement has become much easier and far more accurate.

So why did we write this book? There's certainly no shortage of books on human factors, statistics, experimental design, and usability methods. Some of those books even cover the more common usability metrics. Does a book that focuses entirely on usability metrics even make sense? Obviously, we think so. In our (humble) opinion, this book makes five unique contributions to the realm of usability publications:

- We take a *comprehensive* look at usability metrics. No other books review so many different usability metrics. We provide details on collecting, analyzing, and presenting nearly every type of usability metric you could possibly use.
- This book takes a *practical approach*. We assume you're interested in applying usability metrics as part of your job. We don't waste your time with unnecessary details. We want you to be able to use these metrics easily every day. If you're interested in the theoretical side, we point you to additional resources.
- We provide help in making the *right decisions* about usability metrics. One of the most difficult aspects of a usability professional's job is deciding whether to collect metrics and, if so, which ones to use. We guide you through the decision process so that you find the right metrics for *your* situation.
- We provide many *examples* of how usability metrics have been applied within different organizations and how they have been used to address specific usability questions. We also provide in-depth case studies to help you determine how best to use the information revealed by the usability metrics.
- We present usability metrics that can be used with *any product or technology*. We take a broad view so that these usability metrics can be helpful throughout your career even as technology and products change.

1.1 ORGANIZATION OF THIS BOOK

This book is organized into three main parts. The first one (Chapters 1–3) provides background information needed to get up to speed on usability metrics. This part is intended for those who are less familiar with usability, data analysis, or statistics.

- Chapter 1 provides an *overview* of usability and usability metrics. We define user experience, usability, and different types of usability metrics; discuss the value of measuring the user experience; and dispel some of the common myths about usability metrics.
- Chapter 2 includes *background* information on usability data and some basic statistical concepts. We walk you through a step-by-step process to set up a

usability study using different metrics and provide a guide for performing common statistical procedures related to different usability methods.

- Chapter 3 focuses on *planning a usability study*, including defining participant goals and study goals and choosing the right metrics for a wide variety of situations.

The second part (Chapters 4–9) reviews five general types of usability metrics, as well as some special topics that don’t fall neatly into any single type. For each metric, we explain what it is, when to use it, and when not to use it. We show you how to collect the data and different ways to analyze and present it. We provide examples of how it has been used in real-world usability studies.

- Chapter 4 covers various *types of performance metrics*, including task success, time on task, errors, efficiency, and ease of learning. These metrics are grouped under an “umbrella” of performance because they measure different aspects of the user’s behavior.
- Chapter 5 looks at *measuring usability issues*. Usability issues can easily be quantified by measuring the frequency, severity, and type of issue. We also discuss some of the debates about appropriate sample sizes and how to capture usability issues reliably.
- Chapter 6 focuses on *self-reported metrics*, such as satisfaction, expectations, ease-of-use ratings, usefulness, and awareness. Self-reported metrics are based on what users share about their experiences, not what the usability specialist measures about their actual behaviors.
- Chapter 7 is devoted to *behavioral and physiological metrics*. These metrics include eye-tracking, facial expressions, and various measures of stress. All of these metrics capture something about how the body behaves as a result of the experience of working with a user interface.
- Chapter 8 discusses *how to combine different types of metrics and derive new metrics*. Sometimes it’s helpful to get an overall assessment of the usability of any product. This global assessment is achieved by combining different types of metrics into a single usability score, summarizing them in a usability scorecard, or comparing them to expert performance.
- Chapter 9 presents *special topics* that we believe are important but that don’t fit squarely into one of the five general categories. These include A/B testing on a live website, card-sorting data, Six Sigma, accessibility data, and return on investment (ROI).

The third part (Chapters 10–11) shows how usability metrics are put into practice. In this part, we highlight how usability metrics are actually used within different types of organizations and how to promote the use of metrics within an organization.

- Chapter 10 presents six *case studies*. Each case study reviews how different types of usability metrics were used, how the data were collected and analyzed, and the results. These case studies were drawn from usability practitioners in various types of organizations, including consulting, government, industry, and not-for-profit/education.
- Chapter 11 provides ten *steps to help you move forward in using metrics* within your organization. We discuss how usability metrics can fit within different types of organizations, practical tips for making metrics work within your organization, and recipes for success.

1.2 WHAT IS USABILITY?

Before we try to measure usability, we should know what it is and what it isn't. There are many definitions of usability—maybe even one for every person involved in the field! We're going to focus on three definitions.

The International Standards Organization (ISO 9241-11) identifies three aspects of usability, defining it as “the extent to which a product can be used by specified users to achieve specified goals with *effectiveness*, *efficiency*, and *satisfaction* in a specified context of use.”

The Usability Professionals Association (UPA) definition focuses more on the product development process: “Usability is an approach to product development that incorporates direct user feedback throughout the development cycle in order to reduce costs and create products and tools that meet user needs.”

In his popular book *Don't Make Me Think*, Steve Krug (2000) provides a simple perspective: “Usability really just means making sure that something works well: that a person of average (or even below average) ability and experience can use the thing—whether it's a website, a fighter jet, or a revolving door—for its intended purpose without getting hopelessly frustrated.”

All three of these definitions, as well as other definitions of usability, share some common themes:

- A *user* is involved.
- That user is *doing* something.
- That user is doing something with a *product, system, or other thing*.

Some people distinguish between the terms *usability* and *user experience*. *Usability* is usually considered the ability of the user to use the thing to carry out a task successfully, whereas *user experience* takes a broader view, looking at the individual's entire interaction with the thing, as well as the thoughts, feelings, and perceptions that result from that interaction. We take a very broad view of usability and examine the entire user experience. Therefore, when we talk about “measuring usability,” we're really looking at the entire user experience.

1.3 WHY DOES USABILITY MATTER?

In any casual conversation about usability, most people would agree that it's good to have something that works well and isn't confusing to use. In our years of evaluating products with thousands of test participants, no one has ever complained that something was too easy to use! Everyone has some favorite stories about how something works remarkably well or really terribly. Even as we write this book, we're challenged in formatting the manuscript to give it to the publisher in an acceptable format. Many other stories demonstrate that the usability of something can actually save lives, bankrupt businesses, and have a tremendous impact on society at large.

Usability can sometimes mean the difference between life and death. For example, the health industry is not immune to poor usability. Usability issues abound in medical devices, procedures, and even diagnostic tools. Jakob Nielsen (2005) cites one study that found 22 separate usability issues that contributed to patients receiving the wrong medicine. Similar situations can arise on a regular basis in the workplace or in the home. Just think of the written instructions for such actions as lighting the pilot light on a furnace or installing a new lighting fixture. An instruction that's misunderstood or misread can easily result in property damage, personal injury, or even death. Similar situations arise outside the home: confusing street signs, complicated or poorly designed automobile dashboards, or distracting devices such as vehicle navigation systems or even cell phones.

Sadly, one of the factors involved in a fatal accident on March 2, 2007, in Atlanta, Georgia, may have been poor usability of the high-occupancy vehicle (HOV) lane and its associated signage (see Figure 1.1). A charter bus was carrying the Bluffton University baseball team to their first playoff game in Sarasota, Florida. The bus was driving south in the predawn hours (when it was still dark) on I-75 in the HOV lane, when the driver, who was not from the area, was faced with the signs in Figure 1.1.

Knowing that he wanted to continue in the HOV lane, the driver had to make a split-second decision about which way to go. At least in the United States, the diamond symbol is a standard symbol for an HOV lane, and it can even be seen marking the HOV lane pavement in the figure. Unfortunately, the driver decided to follow the sign that marked the path that branched to the left. The intent of the sign designer was to mark an *exit* from the HOV lane, not the continuation of the HOV lane itself. The bus reached the end of the ramp, continuing at normal speed, when the driver suddenly saw a stop sign and an abrupt end of the road at an overpass. The driver was unable to stop the bus, and it careened off the other side of the overpass and crashed back onto I-75. Seven people died, including the bus driver. If you had been assigned the task of evaluating possible signage for this exit, how might you have done that? What usability metrics could you have used?

Saving lives is, of course, not the only motivation for good usability. Championing usability in a business setting is often geared toward increasing revenues and/or decreasing costs. Stories abound of companies that lost money because of the poor usability of a new product. Other companies have made ease of use a key



FIGURE 1.1

Signs at the time of the March 2, 2007, bus accident in Atlanta, GA. The “Northside Drive” sign, which marks an exit, may have been interpreted as a continuation of the HOV lane.

selling point and have profited by this approach. Staples, the office supply store, is an excellent example of a company that has embraced the concept of ease of use, even in their advertising. Other companies are also beginning to promote ease of use in their product or brand.

Usability can have a tremendous impact on our society. Consider the infamous “butterfly ballot” used in Florida during the 2000 U.S. presidential election. As a result of the usability problems encountered in that ballot, the Usability Professionals Association started a “design for democracy” group that looks at such things as ballot design, voting machines, and the usability of absentee ballots. Outside the political arena, usability can have a huge impact on providing access to goods and services for different user populations, such as older adults, people with disabilities, or people with language or literacy challenges.

Usability plays a much wider role in our lives than most people realize. It’s not just about using a website, a piece of software, or the latest technology. Usability is about setting up a tent, relighting a furnace to heat a home, trying to figure out a tax form, or driving an unfamiliar rental car. Usability impacts everyone, every day. It cuts across cultures, age, gender, and economic class.

Usability takes on an ever-increasing role in our lives as products become more complex. As technologies evolve and mature, they tend to be used by an increasingly diverse set of users. For example, consider the evolution of the written word.

In the earliest days, books were extremely rare (and beautiful) things that were written and illustrated by monks over many years. Precious few people had access to these books or could read them. The invention of the printing press by Johann Gutenberg in 1450 changed that. Books became more widely available, and over many generations more people learned to read. The introduction of electronic books, audio books, and the web brought even more impressive changes. Now even someone who is blind can “read” a book using modern technology.

But this kind of increasing complexity and evolution of technology doesn't necessarily mean that the technologies are becoming easier to use. In fact, just the opposite is likely to happen unless we pay close attention to the user experience. Anyone who's tried to set the time on any of a variety of household devices flashing “12:00” after a power failure can attest to that! As the complexity of technology grows, we believe that usability must be given more attention and importance, and usability metrics will become a critical part of the development process to provide complex technology that's easy to use. We believe that the design process will become more and more user centered.

1.4 WHAT ARE USABILITY METRICS?

A *metric* is a way of measuring or evaluating a particular phenomenon or thing. We can say something is longer, taller, or faster because we are able to measure or quantify some attribute of it, such as distance, height, or speed. The process requires agreement on how to measure these things, as well as a consistent and reliable way of doing it. An inch is the same length regardless of who is measuring it, and a second lasts for the same amount of time no matter what the time-keeping device is. Standards for such measures are defined by a society as a whole and are based on standard definitions of each measure.

Metrics exist in many areas of our lives. We're familiar with many metrics, such as time, distance, weight, height, speed, temperature, volume, and so on. Every industry, activity, and culture has its own set of metrics. For example, the auto industry is interested in the horsepower of a car, its gas mileage, and the cost of materials. The computer industry is concerned with processor speed, memory size, and power requirements. At home, we're interested in similar measurements: the bathroom scale (losing or gaining weight), the curfew for our children, and the size of the turkey required to feed everyone on Thanksgiving.

The usability field is no different. We have a set of metrics specific to our profession: task success, user satisfaction, and errors, among others. This book gathers all the usability metrics in one place and explains how to use these metrics to provide maximum benefit to you and your organization.

So what is a usability metric and how does it compare to other types of metrics? Like all other metrics, usability metrics are based on a reliable system of measurement: Using the same set of measurements each time something is measured should result in comparable outcomes. All usability metrics must be *observable*

in some way, either directly or indirectly. This observation might be simply noting that a task was completed successfully or noting the time required to complete the task. All usability metrics must be *quantifiable*—they have to be turned into a number or counted in some way. All usability metrics also require that the thing being measured represent some aspect of the user experience, presented in a numeric format. For example, a usability metric might reveal that 65 percent of the users are satisfied with using a product, or that 90 percent of the users are able to complete a set of tasks in less than one minute.

What makes a usability metric different from other metrics? Usability metrics reveal something about the user experience—about the personal experience of the human being using the thing. A usability metric reveals something about the interaction between the user and the thing: some aspect of *effectiveness* (being able to complete a task), *efficiency* (the amount of effort required to complete the task), or *satisfaction* (the degree to which the user was happy with his or her experience while performing the task).

Another difference between usability metrics and other metrics is that they measure something about *people* and their behavior or attitudes. Because people are amazingly diverse and adaptable, we sometimes encounter challenges in our usability metrics. For this reason, we will discuss *confidence intervals* with most of the usability metrics we discuss.

Certain things are not considered usability metrics, such as overall preferences and attitudes not tied to an actual experience of using something. Think of some standard metrics such as the Presidential Approval Ratings, the Consumer Price Index, or the frequency of purchasing specific products. Although these metrics are all quantifiable and may reflect some type of behavior, they are not based on actually using something in order to reflect the variability in the data.

Usability metrics are not an end unto themselves; rather, they are a means to help you reach an informed decision. Usability metrics provide answers to questions that are critical to your organization and that can't be answered by other means. For example, usability metrics can answer these critical questions:

- Will the users like the product?
- Is this new product more efficient to use than the current product?
- How does the usability of this product compare to the competition?
- What are the most significant usability problems with this product?
- Are improvements being made from one design iteration to the next?

1.5 THE VALUE OF USABILITY METRICS

We think usability metrics are pretty amazing. Measuring the user experience offers so much more than just simple observation. Metrics add structure to the design and evaluation process, give insight into the findings, and provide information to the decision makers. Without the information provided by usability metrics, important

business decisions are made based on incorrect assumptions, “gut feelings,” or hunches. As a result, some of these decisions are not the best ones.

During a typical usability evaluation, it’s fairly easy to spot some of the more obvious usability issues. But it’s much harder to estimate the size or magnitude of the issues. For example, if all eight participants in a study have the same exact problem, you can be quite certain it is a common problem. But what if only two or three of the eight participants encounter the problem? What does that mean for the larger population of users? Usability metrics offer a way to estimate the number of users likely to experience this problem. Knowing the magnitude of the problem could mean the difference between delaying a major product launch and simply adding an additional item to the bug list with a low priority. Without usability metrics, the magnitude of the problem is just a guess.

Usability metrics show whether you’re actually improving the user experience from one product to the next. An astute manager will want to know as close to certain as possible that the new product will actually be better than the current product. Usability metrics are the only way to really know if the desired improvements have been realized. By measuring and comparing the current with new, “improved” product and evaluating the potential improvement, you create a win-win situation. There are three possible outcomes:

- The new version tests better than the current product: Everyone can sleep well at night knowing that improvements were made.
- The new version tests worse than the current version: Steps can be taken to address the problem or put remediation plans into place.
- No difference between the current product and the new product is apparent: The impact on the user experience does not affect the success or failure of the new product. However, improvements in other aspects of the product could make up for the lack of improvement in the user experience.

Usability metrics are a key ingredient in calculating a ROI. As part of a business plan, you may be asked to determine how much money is saved or how revenue increases as a result of a new product design. Without usability metrics, this task is impossible. With usability metrics, you might determine that a simple change in a data input field on an internal website could reduce data entry errors by 75 percent, reduce the time required to complete the customer service task, increase the number of transactions processed each day, reduce the backlog in customer orders, cut the delay in customer shipments, and increase both customer satisfaction and customer orders, resulting in an overall rise in revenue for the company.

Usability metrics can help reveal patterns that are hard or even impossible to see. Evaluating a product with a very small sample size (without collecting any metrics) usually reveals the most obvious usability problems. However, there are many more subtle problems that require the power of metrics. For example, sometimes it’s difficult to see small inefficiencies, such as the need to reenter user data whenever a transaction displays a new screen. Users may be able to

complete their tasks—and maybe even say they like it—but a bunch of small inefficiencies can eventually build up to impact the user experience and slow down the process. Usability metrics help you gain new insights and lead toward a better understanding of user behavior.

1.6 TEN COMMON MYTHS ABOUT USABILITY METRICS

There are many common myths about usability metrics. Some of these myths may come from a lack of experience with using metrics. Perhaps these myths arose from a negative experience (such as someone from marketing screaming about your sample size) or even other usability professionals complaining about the hassles and costs associated with using metrics. Ultimately the source of these myths doesn't matter. What matters is to separate fact from fiction. We've listed ten of the most common myths surrounding usability metrics and a few examples that dispel these myths.

Myth 1: Metrics take too much time to collect

At best, usability metrics can speed up the design process and, at worst, should not impact the overall timeline. Metrics are quickly and easily collected as part of normal iterative usability evaluation. Project team members may incorrectly assume that full-blown surveys need to be launched or that you have to be testing in the lab for two straight weeks to collect even basic usability metrics. In fact, there are some fairly simple usability metrics you can collect as part of your everyday testing. Adding a few extra questions at the beginning or end of each usability session will not impact the length of the session. Participants can quickly answer a few key questions as part of either a typical background questionnaire or follow-up activities.

Participants can also rate tasks for ease of use or satisfaction after each task or at the end of all tasks. If you have easy access to a large group of target users or a user panel, you can send out an e-mail blast with a few key questions, perhaps with some screenshots. It's possible to collect data from hundreds of users in just one day. Some data can also be quickly collected without even involving the user. For example, you can quickly and easily report the frequency and severity of specific issues with each new design iteration. The time it takes to collect metrics doesn't have to be weeks or even days. Sometimes it's just a few extra hours or even minutes.

Myth 2: Usability metrics cost too much money

Some people believe that the only way to get reliable usability data is to outsource the study to a market research firm or usability consultancy. Although this may be helpful in some situations, it can also be quite costly. Many reliable metrics don't

cost an arm and a leg. Even as part of your everyday testing, you can collect incredibly valuable data on the frequency and severity of different usability issues. You can also collect huge amounts of quantitative data by sending out short e-mail surveys to fellow employees or a panel of targeted users. Also, some of the best analysis tools are actually free on the web. Although money does help in certain situations, it is by no means necessary to get some great metrics.

Myth 3: Usability metrics are not useful when focusing on small improvements

Some project team members may question the usefulness of metrics when they are interested in only some fairly small improvements. They may say it's best to focus on a narrow set of improvements and not worry about metrics. They may not have any extra time or budget to collect any usability metrics. They may say that metrics have no place in a rapid-pace iterative design process. Analyzing usability issues is an obvious and incredibly valuable solution. For example, looking at the severity and frequency of usability issues and why they occur is an excellent way to focus resources during the design process. This approach saves the project both money and time. You can easily derive usability metrics based on previous studies that might help you answer key usability questions. Usability metrics are useful for large and small projects alike.

Myth 4: Usability metrics don't help us understand causes

Some people argue that metrics don't help us understand the root cause of usability problems. They assume (incorrectly) that metrics serve only to highlight the magnitude of the problem. But if they concentrate on only success rates or completion time data, it's easy to see why some might have this perception. Metrics, however, can tell you much more about the root cause of usability issues than you might initially think. You can analyze verbatim comments to reveal the source of the problem and how many users experience it. You can identify where in the system users experience a problem and use metrics to tell where and even why some problems occur. Depending on how the data are coded and the methods used, there is a wealth of usability data that can help reveal the root cause of many usability issues.

Myth 5: Usability data are too noisy

One big criticism of usability metrics is that the data are too “noisy”: Too many variables prevent getting a clear picture of what's going on. The classic example of “noisy” data is measuring task completion time in an automated usability study when the participant goes out for a cup of coffee or, worse, home for the weekend. Although this may happen on occasion, it should not deter you from collecting task time data or any other type of usability data. There are some simple things

that can be done to minimize or even remove noise in the data. Usability data can be cleaned up so that extreme values are not used in the analysis. Also, specific metrics can be carefully chosen to mitigate noisy data. Well-defined procedures can be used to ensure that appropriate levels of consistency are achieved in evaluating tasks or usability issues. Many standard usability questionnaires have already been widely validated by many researchers. The bottom line is that with some careful thought and a few simple techniques, a lot of the noise in usability data can be significantly reduced to show a clear picture of user behavior and attitudes.

Myth 6: You can just trust your gut

A lot of usability decisions are made on a “gut level.” There’s always someone on the project team who proclaims, “This decision just feels right!” One of the beauties of metrics is that having the data takes a lot of the guesswork out of usability decisions. Some design options are truly borderline cases, but they might actually have an impact on a large population. Sometimes the right design solutions are counterintuitive. For example, a design team may ensure that all the information on a web page is above the fold, thereby eliminating the need to scroll. However, usability data (perhaps in the form of task completion times) may reveal longer task completion times because there’s not enough white space between the various visual elements. Intuition is certainly important, but data are better.

Myth 7: Metrics don’t apply to new products

Some people shy away from metrics when evaluating a new product. They may argue that since there is no point of comparison, metrics don’t make sense. We would argue just the opposite. When evaluating a new product, it’s critical to establish a set of baseline metrics against which future design iterations can be compared. It’s the only way to really know if the design is improving or not. In addition, it’s helpful to establish target metrics for new products. Before a product is released, it should meet basic usability metrics around task success, satisfaction, and efficiency.

Myth 8: No metrics exist for the type of issues we are dealing with

Some people believe that there aren’t any metrics related to the particular product or project they are working on. Whatever the goal of the project, at least a couple of metrics should tie directly to the business goals of the project. For example, some people say they are only interested in the emotional response of users and not in actual task performance. In this case, several well-established ways of measuring emotional responses are available. In other situations, someone might be concerned only with awareness. Very simple ways to measure awareness also exist, even without investing in eye-tracking equipment. Some people say that they are only interested in more subtle reactions of users, such as their level of frustration. There

are ways to measure stress levels without actually asking the user. In our years of usability research, we have yet to come across a business or user goal that was not measurable in some way. You may have to be creative in how you collect the data, but it's always possible.

Myth 9: Metrics are not understood or appreciated by management

Although some managers view usability as providing only qualitative feedback about a design or product, most managers see the value of measurement. It has been our experience that usability metrics are not only understood but very much appreciated by upper-level management. They can relate to metrics. Metrics provide credibility to the team, the product, and the design process. Metrics can be used to calculate ROI (return on investment). Most managers love metrics, and usability metrics are one type of metric they will quickly embrace. Usability metrics can also be real attention-grabbers with management. It's one thing to say there's a problem with the online checkout process, but it's an entirely different thing to say that 52 percent of users are unable to successfully purchase a product online once they've found it.

Myth 10: It's difficult to collect reliable data with a small sample size

A widely held belief is that a large sample size is required to collect any reliable usability metrics. Many people assume that you need at least 30 participants to even start looking at usability data. Although having a larger sample size certainly helps increase the confidence level, smaller sample sizes of eight or ten participants can still be meaningful. We will show you how to calculate a confidence interval that takes into account the sample size when making any conclusion. Also, we will show you how to determine the sample size you need to identify usability issues. Most of the examples in this book are based on fairly small sample sizes (fewer than 20 participants). So not only are metrics possible to analyze with fairly small sample sizes, doing so is quite common!