Special Issue

Usability: Adoption, Measurement, Value

Janine D. Mator, **William E. Lehman**, **Wyatt McManus**, **Sarah Powers**, **Lauren Tiller**, **James R. Unverricht**, and **Jeremiah D. Still**, Old Dominion University, Norfolk, VA, USA

Objective: We searched for the application of usability in the literature with a focus on adoption, measurements employed, and demonstrated value. Five human factors domains served as a platform for our reflection, which included the last 20 years.

Background: As usability studies continue to accumulate, there has been only a little past reflection on usability and contributions across a variety of applications. Our research provides a background for general usability, and we target specific usability research subareas within transportation, aging populations, autistic populations, telehealth, and cybersecurity.

Method: "Usability" research was explored across five different domains within human factors. The goal was not to perform an exhaustive review but, rather, sample usability practices within several specific subareas. We focused on answering three questions: How was usability adopted? How was it measured? How was it framed in terms of value?

Conclusion: We found that usability is very domain specific. Usability benchmarking studies and empirical standards are rare. The value associated with improving usability ranged widely—from monetary benefits to saving lives. Thus, researchers are motivated to further improve usability practices. A number of data collection and interpretation challenges still call for solutions.

Application: Findings offer insight into the development of usability, as applied across a variety of subdomains. Our reflection ought to inform future theory development efforts. We are concerned about the lack of established benchmarks, which can help ground data interpretation. Future research should address this gap in the literature. We note that our findings can be used to develop better training materials for future usability researchers.

Keywords: human–computer interaction, interface evaluation, computer interface, product design

Address correspondence to Jeremiah D. Still, Department of Psychology, Old Dominion University, Mills Godwin Building 236, Norfolk, VA 23529, USA; e-mail: jstill@odu.edu

HUMAN FACTORS
Vol. 00, No. 0, Month XXXX, pp. 1
DOI:10.1177/0018720819895098
Article reuse guidelines: sagepub.com/journals-permissions

INTRODUCTION

While the birth of usability is commonly associated with the 1980s, its history may be traced to decades prior. The field of human factors psychology as a whole arguably emerged during the 1940s, as the military concerns of World War II drove a need to understand the factors of cockpit design that were responsible for pilot error (Stuster, 2006). Through questions of soldier training and precision, human factors research gave way to more concrete usability metrics such as effectiveness (e.g., improving target acquisition by improving the design of the cannon) and efficiency (e.g., considering crew experience to maximize rounds fired per minute) (Soegaard, 2012). Metrics of user satisfaction were noticeably absent, which is unsurprising given the priority of military error reduction. Years later, however, usability would come to be defined as "the extent to which a product can be used with effectiveness, efficiency, and satisfaction in a specified context of use" (International Organization for Standardization, 1998). This industry standard refers to *outcomes* of usability, rather than a single feature of a system, product, or service. More specifically, effectiveness can be measured by accuracy and completeness of specific goals; efficiency can be measured by the cost of resources (such as time and effort) to attain the results; and satisfaction can be measured by how well users' physical, cognitive, and emotional responses meet their own needs and expectations.

Of course, the origin of usability is more commonly associated with the 1980s due to the boom of the personal computer (PC). Although PCs became more affordable and more accessible to the average consumer, their intricate software still catered to users with more technical vocabulary. As a result, usability emerged to design for a broader audience—including for computer novices (Cockton, 2013).

From the late 1980s until the early 1990s, usability specialists seemingly became the new jacks-of-all-trades. Their responsibilities often lacked definition and ranged from user interface design to usability requirements analysis (Mayhew, 2008). However, publications did not explicitly refer to usability until the late 1980s, with the work of John Whiteside and John Bennett at Digital Equipment Corporation and at IBM, respectively. With their publications came the emphasis on the kind of iterative evaluations, integrated product design teams, and cost–benefit analyses commonly used today (Dumas, 2007).

Initially, the approach to usability continued through a traditional cognitive perspective, with causal relationships between design features and user behavior being the goal (Cockton, 2013). During this early phase, usability testing was still conducted as a kind of traditional research experiment. In 1986, the US Air Force commissioned Smith and Mosier to develop its comprehensive "Guidelines for Designing User Interface Software" (Cockton, 2013). However, it was soon realized that the sheer variability of interface features, from icons to window arrangements, did not lend itself well to traditional, comprehensive research experiments (Grudin, 2008).

The work of Bennett and Whiteside in 1988, as previously mentioned, also offered a solution to the challenge of obtaining empirical measurements. Says Dumas (2007), "Instead of stressing the research experiment, they stressed a quantitative but practical engineering approach to product design. The approach stressed early goal setting, prototyping, and iterative evaluation – the foundations of our development methods." Not surprisingly, a wealth of usability evaluation methods ensued. The 1990s saw Polson, Lewis, Rieman, and Wharton's proposal of cognitive walkthroughs in 1991, as well as a variety of new usability questionnaires, such as the Post-Study System Usability Questionnaire, or PSSUQ, and the After-Scenario Questionnaire, or ASQ (Lewis, 1992, 1995). Overall, heuristic evaluation became the most prominent method used in the 1990s. Jacob Nielsen's "10 Usability Heuristics for User Interface Design" stands as one prominent example (Cockton, 2013; Nielsen, 1995).

The prevalence of heuristic evaluation highlights the formative phase of usability. Rather than conducting experiments that strictly adhere to the scientific method, the focus has shifted to the identification of design flaws through quick and dirty testing with minimal controls. Such testing is characterized by collecting small sample sizes and helping participants with midtask assistance (Jordan et al., 2014; Lazar et al., 2010; Wixon, 2003). Industry standards, in addition to monetary benefits, have catalyzed the empirical growth of usability. In 1981, industry standards emerged, and they were quickly followed by an American National Standards Institute (ANSI) group for office and text systems (Grudin, 2008). More recently, the last 10 years have seen an influx of international standards for user-centered design. Similarly, these standards for interfaces in HCI are dominated by general guiding principles, rather than by precise specifications (Bevan, 2001).

Usability and Human Performance

Successful technology developers consider product usability engineering costs critically important (Bias & Mayhew, 1994; Lund, 2011). Good interface usability facilitates human performance (e.g., decreases help calls) and increases market share. However, determining whether a product is usable can be difficult given that it is not possible to directly measure usability (Hornbæk, 2006). We can only create usability constructs that reflect our choice of measurement (Table 1). The construct of perceived usability often emerges from three key contributors: the user, the surrounding environment, and the product design (Johnson, 2008; Norman, 2004). A change in one of these contributors may impact key usability measurements. For example, the product might transform from being viewed as wonderful to being seen as only acceptable because of a change in user type. We often think about the interface's users first when considering the usability of a product; however, depending on one's domain, the

TABLE 1: Prevalence and Measures of Efficiency, Effectiveness, and Satisfaction

Domain	Order of Prevalence	Measures of Efficiency	Measures of Effectiveness	Measures of Satisfaction	Other Measures
Transportation (in-vehicle information systems)	S, ES, EY	Task completion time (e.g., comparison to 15 s rule)	Task completion time	Subjective questionnaires	Eye fixations, driving performance (e.g., steering and speed variation)
Aging populations	S, ES, EY	Error rate, task completion time	Task completion rate, success rate	Acceptability, general opinions of use	Usefulness, learnability, affect state, workload
Autistic populations	S, EY, ES	Number of clicks	Task completion time	Appeal of interfaces, adequacy of help/ documentation, quality of feedback and error messages	Learnability
Telehealth (EHR systems)	EY, ES, S	Number of clicks, number of requests for help	Task completion time, proportion of tasks completed	Ease, acceptance, amount of time, quality of support information	N/A
Cybersecurity (graphical authentication)	EY, ES, S	Login times	Login success	Subjective questionnaires	Learnability, memorability

Note. The three main constructs of usability (EY for efficiency, ES for effectiveness, and S for satisfaction) are listed in order of prevalence among usability studies for each domain. Specific measures are stated for each construct. Notably, completion time was used to measure both effectiveness and efficiency across studies depending on the nature of the task. EHR = electronic health record.

focus might be on other usability contributors. Inherent conflicts are natural when considering interface design trade-offs (Lazar et al., 2010). For example, cybersecurity might be more focused on environmental factors (e.g., maintaining security, stakeholder requirements) than on a user's cognitive abilities or preferences.

Usability research needs to create meaningful standards. This can be challenging, given the diverse use of usability. We need to explore what measures are currently being employed and whether standards are being utilized. Benchmarking enables designers to recognize and achieve these standards, also known as best practices, for their products and interfaces (Bhutta & Huq, 1999). With the development of similar systems and prototypes, benchmarking data makes it possible to draw meaning from cross-study comparisons and determine how well one design performs compared to another (Bhutta & Huq, 1999).

Benchmarks vary widely across domains—for example, task completion time standards may be 15 s for in-vehicle information systems, whereas login times for authentication schemes may be designed with a goal of 7–20 s (Braz & Robert, 2006; Green, 1999). Accuracy and task completion rate are a few other benchmark metrics for assessing the objective usability of a project, while subjective measures may include self-reports of perceived usability, affect, and workload (MacDonald & Atwood, 2013; Sonderegger et al., 2016). According to Eklund and Levingston (2008), usability benchmarks seek to capture two things. First is user needs, requirements, and perceptions, as demonstrated through "what users say and do." Second is expert analysis, demonstrated through performance on specific criteria based on heuristics and guiding principles such as Neilsen's "10 Usability Heuristics" (1995).

More traditional human factors fields have employed similar benchmarks as well. For example, usability assessment in aviation must meet certain expectations of operator workload to ensure that aircraft controls will fall within an acceptable range. Benchmarking is also used to evaluate the performance of simulated pilot tasks compared to actual pilot tasks, similar to the use of modern driving simulators (Clamann & Kaber, 2004; Feng et al., 2017). Ultimately, no matter what the performance measurement, the intent of usability is to provide an improved design. This intent may be realized through an assessment with a focus on human performance, but in addition to performance metrics, usability researchers must also apply usability as a process to develop more general testing protocols (Booker et al., 2004; Rubin & Chisnell, 2008). To this end, we apply usability more broadly throughout this paper to inform design theory.

Usability Across a Variety of Domains

Human-centered design and usability go hand in hand in developing useful products. Interestingly, these are usually taught as generic processes and as assessment tools void of any specific domain details. This has led to the blind application of usability processes, which can be harmful or wasteful (Greenberg & Buxton, 2008). However, human factors psychology is typically domain specific, and the goal of this literature review is not to provide an exhaustive account. Rather, we aim to provide a broad, qualitative overview of usability across a variety of domains. We seek to reflect on three questions: 1. How was usability adopted given unique domains? 2. How was usability measured? 3. How was the value of usability framed?

Systematic Review

We examined the application of usability within five different human factors domains. In order to maintain a digestible scope, a specific research area was identified within each domain representing a specific research area for each coauthor. We explored in-vehicle systems; aging populations; autistic populations; telehealth electronic records; and cybersecurity graphical authentication. These literature searches were not meant to be exhaustive but, rather, to be representative of recent contributions.

To be included in the literature search, the source must have appeared between 1999 and 2019. Critically, the research must be centered on usability within a targeted domain area.

Literature reviews were conducted using the following search engines: Google Scholar, the IEEE *Xplore* Digital Library, ACM Digital Library, PsycINFO, ScienceDirect, and EBSCOhost. Searches included a forward and backward citation examination.

General search terms included usability, usability testing, usability challenges, usability guidelines, and user interface (UI). Domain-specific search terms included variations of general search terms (e.g., "usability in transportation" or "vehicle system usability"), as well as other relevant search terms.

From 5 individual papers representing each domain, 89 of the 128 references were used to create an integrative review. References were selected based on relevance to the three research questions, excluding more tangential information. In addition to meeting specific publication dates and relevant key phrases, articles needed to meet other inclusion criteria: including usability testing as the primary method of system analysis; pertaining to a specific product within the larger domain (e.g., transportation); including a usability measure; including either qualitative or quantitative data involving the development or testing of technology; including a specific subset of the user population (e.g., testing with individuals diagnosed with autism spectrum disorder (ASD), rather than their teachers or parents); contributing a full-length original research paper; and describing how each usability construct was measured.

Sources included books, proceedings, technical reports, literature reviews, and a small number of websites, in addition to peer-reviewed journals.

HOW WAS USABILITY ADOPTED WITHIN A UNIQUE DOMAIN?

Transportation: In-Vehicle Information Systems

According to the Bureau of Transportation Statistics, the number of registered vehicles on the road in the United States has increased from approximately 74 million in 1960 to over 268 million in 2017. As technology continues to advance in the automotive industry, in-vehicle information systems are becoming more prevalent (Ma & Du, 2017). "In-vehicle information system" is the encompassing term for systems designed to enhance the driving experience of the user by providing them with additional information, whether it be entertainment, navigation, or some other function (Mitsopoulos-Rubens et al., 2011). However, any function of the in-vehicle information systems is secondary to driving itself.

The increased use of in-vehicle information systems in passenger vehicles has fostered concern about how these systems may distract the driver from the primary task of driving (Mitsopoulos-Rubens et al., 2011). Empirical studies have found that distraction degrades driving performance, making it more difficult for drivers to keep their vehicles in their own lane and to brake in time to avoid hazardous events (Horrey & Wickens, 2004; Peng et al., 2013). Whereas usability testing often focuses solely on efficiency and performance, using the product itself, the secondary nature of in-vehicle information systems, presents a unique challenge to the domain of transportation (Li et al., 2017).

Usability testing in transportation requires measurement of both primary and secondary task performance (Barón & Green, 2006). Of course, a true measurement of driving while using an in-vehicle information system would require drivers to put themselves at risk by participating in distracted driving. In order to minimize the risk of actual harm to the participant, the use of driving simulators is common when testing the usability of in-vehicle information systems (Feng et al., 2017; Ma & Du, 2017; Mitsopoulos-Rubens et al., 2011; Sumie et al., 1998). Sumie et al. (1998) performed a set of experiments designed to test the usability of hierarchical menu controls for an information system. The overall case for driving simulators has been strong in the past (Godley et al., 2002; Lee et al., 2003), although researchers should consider how actual driving performance might vary, for better or worse, when drivers are facing real-life consequences and unobserved (Reed & Green, 1999). Ultimately, usability testing of in-vehicle information systems informs the design process to maximize driver safety (Stevens et al., 2002).

Aging Populations

Contrary to the commonly held opinion that aged adults are unwilling to use modern technology, research shows that more than half of adults aged 60–70+ use smartphones, laptops, and desktop computers (Anderson, 2016). These computing devices are used for tasks ranging from communication to medical care (Anderson, 2016). Fully 91% of adults aged 50+ reported using technology to stay in touch with family and friends (Anderson, 2016).

The need for usability testing in this domain stems mainly from cognitive weaknesses. For aged adults, the most significant cognitive weakness is fluid intelligence, which is attributed to poor working memory, spatial cognition, and multitasking ability. Decreased fluid intelligence is also reflected by poor prospective memory, as well as by slower learning (Czaja & Lee, 2007; Matthews et al., 2000; Schaie, 1994; Sixsmith, 2013; Ziefle & Bay, 2005). For aging populations, there has been a need to develop usable devices that do not rely heavily on fluid intelligence. Of course, large variability exists between individual abilities (Czaja & Lee, 2007). Age alone should not serve as an indicator of cognitive ability, but rather as a guide to the possible cognitive changes that may occur across the population.

A recently recurring theme within the literature was whether aged adults would accept new technologies. Researchers were fairly split, with seven suggesting acceptance from aged adults (Anderson, 2016; Czaja & Lee, 2007; Hanson & Crayne, 2005; Otjacques et al., 2009; Petrovčič et al., 2018; Sixsmith, 2013;

Ziefle & Bay, 2005) and six either inferring or explicitly stating that aged adults would not be accepting of technology (Akatsu & Miki, 2004; Charness & Boot, 2009; Hernández-Encuentra et al., 2009; Hitchcock et al., 2001; Keith, 2010; Wang et al., 2011). It should also be noted that aged adults may have different acceptance of new technology across domains. For example, an aged adult might be more accepting of using a fall risk application on their smartphone, but would not be willing to relinquish control of a vehicle to artificial intelligence. Depending on the domain, researchers' focus tends to shift from assuming that aged adults would use the technology to offering strategies for implementation, to increase acceptance of new technologies.

Autistic Populations

The number of ASD diagnoses is on the rise. In 2014, around 1 in 59 children were diagnosed with ASD in the United States alone, meaning that usability issues for these users are becoming more prolific (Baio et al., 2018). Technology provides one way to enhance this community's options to interact with the world and to pursue education and work (Bahiss et al., 2010). However, the needs of users with ASD are different from the needs of neurotypical users and those with other intellectual disabilities (Chen et al., 2009). Challenges such as the potential for overstimulation or fixation have resulted in unique design considerations for those with ASD.

First, it is important to note that user variability poses a challenge within this population. Individuals with ASD exhibit wide variability in preferences and abilities related to speech, communication, motor control, and perception (Chen et al., 2009; Fabri et al., 2016; Monibi & Hayes, 2008; Zakari et al., 2014).

Many with ASD face sensory disorders, meaning that users may be especially sensitive to the different levels and types of light or sound present in an interface (Zakari et al., 2014). Loud noises, in particular, tend to upset users with ASD (Barry et al., 2008). For these individuals, it is especially important to reduce the gap between user expectations and system interactions (Barry et al., 2008).

Another need for improved usability in this domain lies in the fact that many individuals with ASD fixate on stimuli that are irrelevant to the task at hand (Wang et al., 2015). This means that users with ASD may become adversely focused on system-irrelevant details. This only reinforces the need to avoid auditory and visual clutter to reduce distraction for these individuals (Artoni et al., 2011).

End users with ASD might also be included in the development and testing process. Participatory design is common, although caution should be exercised so that participants are not made to feel like guinea pigs—especially when they are only enlisted for a single session (Artoni et al., 2010; Bahiss et al., 2010; Brosnan et al., 2016; Cobb et al., 2010; Fabri et al., 2016; Hirano et al., 2010; Millen et al., 2010).

Telehealth: EHR Systems

Electronic health records (EHRs) systems are an important aspect of telehealth. EHRs within the system "facilitate electronic collection and organization of an array of important information, from basic demographics to billing and coding reports to medications to laboratory values" (Miriovsky et al., 2012). ISO 62366 (2008) is a more recent medical equivalence of ISO 9241–11 that mirrors concerns of effectiveness, efficiency, and user satisfaction. However, particular emphasis is placed on reducing risk and error (International Electrotechnical Committee (IEC), 2015).

While EHR systems vary across manufacturers and intended use, there are several usability challenges that have been consistently identified across EHR systems in the literature: universal design, clear navigation, ease of data entry, and proneness to error.

First, EHR systems need to be accessible to all users. Typically, the first step in designing a usable product is to know the abilities and needs of the user population (Rubin & Chisnell, 2008). However, like the autistic population, users who access EHRs are diverse. Representative users consist of both medical staff and patients of various genders, ethnicities, ages, and knowledge.

Singh et al. (2010) found that the usability of EHR systems is often focused on the needs of physicians rather than the needs of the patients. Even within the patient user population, some may struggle more than others. In the same study, older adults were found to have less computer experience than the general population, often resulting in the kind of anxiety associated with using technology. As previously mentioned, fluid intelligence is a significant challenge to usability among aging populations—a challenge that may be exacerbated by poor usability (Matthews et al., 2000; Schaie, 1994).

Second, EHR systems need to be easy to navigate. Navigating EHR system interfaces can be difficult both for young and old patients (Chun & Patterson, 2012). Segall et al. (2011) conducted a usability test that showed that more than half of the participants had a difficult time navigating an EHR system to complete several basic tasks. Furthermore, when interviewed about the usability of the system, navigation was rated by participants as one of the poorest factors. Some reasons given for poor navigation were cluttered displays, lack of a hierarchical menu structure, and lack of intuitive pathways to the intended information (Chun & Patterson, 2012). Several researchers have suggested that the volume of medical data is a contributing factor in this poor navigation. With such a large volume of complex data, clear organization becomes difficult (Goldberg et al., 2011). For EHR systems, a unique factor is that the data is often fragmented across screens, forcing users to hold information in their working memory while trying to piece it all together (Institute of Medicine, 2012). Consequently, users will refuse to use an EHR system with poor navigation because it is effortful. Poorly designed navigation systems result in increased cognitive load, increased time to use the system, and numerous errors, which harms productivity (Smelcer et al., 2009).

Another usability issue of EHR systems found in the literature is the difficulty of data entry. In a usability test conducted by Sarkar et al. (2016), less than half of the participants were able to add an appointment or a medication into the system. Also, issues with data entry are not limited to patients. O'Connell et al. (2004) found that 72% of physicians also found data entry difficult.

A final need prevalent in the literature is reducing the likelihood of committing an error. Using an expert panel, Shneiderman (2011) identified types of errors prominent in EHR systems. Several of these included wrong patient actions, incorrect treatment actions, wrong medication actions, delays in treatment events, and unintended or improper care events. Compounding the problem, Schumacher et al. (2010) discovered that error messages are often poorly designed, resulting in physicians ignoring the messages even when they were critical. Improvements to patient care may, therefore, be aided by improvements to EHR system usability.

Cybersecurity: Graphical Authentication

Authentication schemes are used to confirm the identity of authorized account holders. Users are commonly granted account or system access through knowledge-based information (e.g., a password), by using a physical object (e.g., a swipe card), or by showing some other proof of identity (e.g., a fingerprint) (Cazier & Medlin, 2006; Grassie et al., 2017). The most commonly used form of authentication is the traditional knowledge-based password (Grawemeyer & Johnson, 2011; Leu, 2017). System designers often use this password authentication scheme because it is conventional (Herley & Van Oorschot, 2011).

Traditional passwords are expected to be complex, unique, and memorable, and to remain secret (Hoonakker et al., 2009). For many users, complex and unique passwords are difficult to remember (Yan et al., 2004). In addition, few users are aware of the risks associated with poor authentication practices (Cain et al., 2018). Previous research has noted that the recommended practices for strong password creation can lead users to create unmemorable passwords. Creating a strong password takes effort, which encourages users to invent cognitive workarounds that undermine security (Still et al., 2017).

Usability in cybersecurity has emerged, in part, through a need for alternatives to traditional authentication—specifically, alternatives that are both secure and usable (Cain et al., 2018; Still et al., 2017). To make authentication more memorable, graphical scheme designers take advantage of an individual's natural ability for recognizing visual objects (Paivio, 2013). Memorability is an important quality that must be considered when developing an authentication scheme, since users may need to remember their passcode following long temporal delays between system interactions (Tiller et al., 2019; Wiedenbeck et al., 2005).

The development of graphical authentication schemes is not only about memorability. Designers also consider attack vectors that are meant to overcome system security. An over-the-shoulder attack (OSA) is the most common cybersecurity threat to graphical authentication schemes (Cain et al., 2018). An OSA occurs when an unauthorized casual screen looker steals the user's passcode in a shared space.

Defining the design space in graphical authentication can be done by considering (1) the technical security component of authentication, and (2) the users' cognitive abilities (Herley & Van Oorschot, 2011). For example, authentication is one area where technical security is a greater priority than usability—just as driving performance takes precedence over use of in-vehicle information systems (Inglesant & Sasse, 2010; Li et al., 2017).

HOW WAS USABILITY MEASURED?

Transportation: In-Vehicle Information Systems

The inclusion of driving performance measures is important to understanding the usability of the system (Ma & Du, 2017). A consistent set of driving performance measures is used when testing a large array of in-vehicle information systems. For example, Weinberg et al. (2011) used a single-session driving simulator methodology to compare heads-down display in-vehicle information systems to heads-up display systems. The authors collected measures of user satisfaction, eye fixations, and task completion time, and the driving performance measures captured lateral vehicle position, steering variation, throttle/brake variation, and speed variation. All of these measures have been consistently used across the literature. Further, they are suggested across multiple usability testing protocols (Harvey et al., 2011; Harvey, 2009; Stevens et al., 2002). The use of driving simulators and single-trial usability research has been validated, and measures of driving performance have been developed and implemented (Sumie et al., 1998). Benchmark standards for the efficiency of in-vehicle information systems do exist, such as the 15-s rule for task completion time (Green, 1999; SAE, 1998, 1999). Organizations such as the National Highway Traffic Safety Association have also provided benchmarks for various aspects of in-vehicle information system usability regarding driver workload, response time, and related concepts (Ranney et al., 2011).

Aging Populations

Satisfaction, effectiveness, and efficiency were the three main usability measures employed within the domain of aging populations. Satisfaction was found to be the most prevalent measure across the literature (Akatsu & Miki, 2004; Garcia-Sanjuan et al., 2017; Hernández-Encuentra et al., 2009; Hsieh et al., 2018; McCarthy et al., 2007; Nilsson et al., 2003; Otjacques et al., 2009; Sonderegger et al., 2016; Wang et al., 2011; Ziefle & Bay, 2005). Less prevalent measures consisted of usefulness (Haslwanter et al., 2018), learnability (Garcia-Sanjuan et al., 2017; Petrovčič et al., 2018), affect state, and workload (Sonderegger et al., 2016).

Satisfaction was measured using a variety of methods, ranging from acceptability to general opinions of use. The second most prevalent measures were effectiveness and efficiency. Only one study used the novel approach of measuring usability through workload or affect state (Sonderegger et al., 2016). In the same study, effectiveness was measured as task completion rate, and efficiency was measured as task completion time and error rate. Others, such as Ziefle and Bay (2005), measured effectiveness by success rate, or percentage of successfully solved tasks; efficiency was measured by task completion time and error rate, or the number of detours and returns in a navigation-based task. Relative comparisons, rather than absolute standards, were found in assessing usability for aging populations.

Autistic Populations

Testing with ASD users may be challenging since they can become overstimulated or have low verbal ability (Hirano et al., 2010). Usability is often captured through the use of questionnaires, such as the System Usability Scale (Fabri et al., 2016).

Questionnaires by Khan et al. (2013) used a five-point scale, ranging from “strongly agree” to “strongly disagree,” to measure the appeal of interfaces; adequacy of help and documentation; overall ease of use; quality of feedback and error messages; and learnability. Other studies asked participants if they would use or recommend the technology (Hatfield et al., 2018; Politis et al., 2017; Van Laarhoven et al., 2018).

Open-ended discussion or think-aloud techniques may be done with participants who are able and willing to talk, but who may prefer not to fill out questionnaires (Caro et al., 2017; Cobb et al., 2002; Fabri et al., 2016; Hatfield et al., 2018; MacLeod, 2010; Politis et al., 2017, 2019). However, these options pose difficulty for users with low verbal abilities (Caro et al., 2017; Hirano et al., 2010). For low or nonverbal users, many studies use observation while users interact with technology (Caro et al., 2017; Checkley et al., 2010; Cibrian et al., 2017; Cobb et al., 2010; Hirano et al., 2010). The observed behavior may be directly observed (number of clicks, time to task completion) or may be interpreted by a caretaker who can identify the meaning of specific user behaviors (Caro et al., 2017; Grynszpan et al., 2008).

Other users perform best using a text–picture combination rather than a text-only or text–speech combination (Chen et al., 2009). For participants with less verbal proficiency, pictorial storyboards or other visual tools can also be used to aid communication (Hirano et al., 2010; Millen et al., 2010: Monibi & Hayes, 2008).

Perhaps due to the variability of users with ASD, standards are difficult to achieve within this domain. Individuals with varying levels of severity of ASD may differ wildly. Comparing standards from the general population to this special population is ill-advised.

Telehealth: EHR Systems

Again we see a focus on efficiency, effectiveness, and satisfaction to validate usability in EHR systems. To measure efficiency, researchers rely on objective measures such as the number of clicks needed to complete a task (Rojas & Seckman, 2014) or the number of requests for help (Segall et al., 2011). Objective measures are also often used to measure the effectiveness of EHR systems. Several measures commonly used include the task completion rate (Chun & Patterson, 2012), the proportion of tasks completed (Sarkar et al., 2016), and the number of errors (Segall et al., 2011).

Consequently, satisfaction associated with using an EHR system is typically measured subjectively with questionnaires (Chun & Patterson, 2012; Rojas & Seckman, 2014; Sarkar et al., 2016; Segall et al., 2011). Little could be determined from previous research regarding benchmarking or standards in telehealth.

Cybersecurity: Graphical Authentication

Efficiency and effectiveness were found to be popular measures of usability in cybersecurity. However, learnability and memorability did appear. The exact methods used to acquire these measures varied greatly from study to study.

Still et al. (2017) have provided a formalized set of design guidelines to improve the usability of authentication schemes. The guidelines suggest that designers consider inclusivity, login times, error rates, learnability, memorability, and subjective satisfaction ratings. Most importantly, an ideal authentication scheme for users should be usable and should provide strong security by restricting unauthorized access.

The purpose of authentication is to keep valuable data secure. If a proposed authentication scheme does not provide the users with ample account security, the whole scheme is rendered useless. Many articles have used calculations such as bit strength entropy (Sun et al., 2012; Wiedenbeck et al., 2005), Hartley entropy (Bianchi et al., 2016), or probability of success with a single random guess (Khot et al., 2012).

Designers are now being asked to provide some information regarding the level of security of their proposed scheme.

On the human side, login times often indicate an authentication scheme's efficiency. Ideally, schemes need to provide users with quick access, in order to be considered efficient (Still et al., 2017). The results of our review indicate that the use of recorded login times, in order to establish the usability of the scheme, is very common. Graphical authentication scheme designers should aim to create a scheme that achieves login times comparable to traditional passwords (e.g., 7–20 s) (Braz & Robert, 2006).

Effectiveness is typically measured in terms of login success. It was observed that researchers define the frequency of success rates differently. The variations of the different accuracy measures are as follows: total authentication accuracy (given six login attempts) (Sun et al., 2012), partial passcode accuracy (Chiasson et al., 2007), number of incorrect passcode submissions (Wiedenbeck et al., 2005, 2006), and success rate (given three attempts) (Bianchi et al., 2016; Hayashi et al., 2008). These measurements are often used to compare schemes across the literature. Subtle differences between measures can make direct comparisons misleading.

A few experiments have measured learnability. For example, Cain et al. (2018) recorded the number of correct attempts over time to evaluate the learnability of four different schemes. On the other hand, Khot et al. (2012) evaluated learnability by way of efficiency and measured login time improvements with practice.

Users are typically interacting with a new graphical authentication scheme and are making learnability a reasonable consideration. Most studies of next-gen authentication schemes provide users with training to introduce the new technology, but few studies measure learnability performance.

We also discovered that memorability is being captured in a few studies. Surprisingly, it was not a popular performance metric, given that users often have large temporal gaps between system interactions. Memorability is typically measured during the initial session and after a 1-week delay (Gao et al., 2008; Wiedenbeck et al., 2006; Zangooei et al., 2012). Fewer studies have examined performance across three time points—that is, following the initial session, after 1 week, and after 6 weeks (Hayashi et al., 2008; Sun et al., 2012; Wiedenbeck et al., 2005).

HOW WAS THE VALUE OF USABILITY FRAMED?

Transportation: In-Vehicle Information Systems

In 2015, distracted driving accounted for 10% of all driving-related deaths in the United States (National Center for Statistics and Analysis, 2017). The issue of distracted driving caused by the inclusion of in-vehicle information systems adds additional usability concerns that must be addressed.

Usability testing of in-vehicle information systems is imperative to ensure the safety of both the user and those on the roadway around them. Only through testing during the development stage can steps be taken to reduce interaction time with the systems—and therefore ensure the user's ability to command the vehicle safely.

As previously mentioned, distracted driving is a major cause of injury and death in the United States (National Center for Statistics and Analysis, 2017), so a major focus of the usability testing for in-vehicle information systems should be the minimization of time required to use its various functions. Seconds spent not attending to the roadway are seconds that drastically increase a user's likelihood to crash (Klauer et al., 2006).

Aging Populations

Improvements to usability for aging populations are readily apparent from related literature on mobile device data. As found by Oksman (2006) in a sample of adults in Finland over the age of 60, negative attitudes toward mobile phones have reversed since 2002. For example, aged adults were less inclined to associate mobile phones with special circumstances. Overall, Oksman reported that mobile communication has "more than fulfilled the expectations directed towards it by the age group of seniors."

It is worth noting that older adults comprise the fastest growing group in the world population (Rubin & Chisnell, 2008). Among the elderly, safety and mobility are even more highly regarded than privacy (Melander-Wikman et al., 2007). To this end, usability of smart home technology is another consideration. Those who are hearing, visually, and/or cognitive impaired may benefit greatly from smart home technology and from virtual assistants. The needs and issues of the elderly extend beyond safety measures as well; social isolation and a loss of independence are frequent concerns, among others (Cheek et al., 2005).

Using these technologies can improve the quality of life for aged individuals. Specifically, it can improve mobility (Hitchcock et al., 2001), foster socialization (Garcia-Sanjuan et al., 2017; Otjacques et al., 2009), and mitigate cognitive declines such as decrements in memory performance or manual skills (Garcia-Sanjuan et al., 2017; Hitchcock et al., 2001; Jacko et al., 2000; McCarthy et al., 2007; Pak et al., 2008). Also, the implementation of technology can save the lives of the aged by providing health information (Hsieh et al., 2018).

Autistic Populations

Previous research has explored education (Caro et al., 2017; Hatfield et al., 2018, Van Laarhoven et al., 2018), social interaction training (Jeon et al., 2015; Millen et al., 2010), social media (Bahiss et al., 2010), serious games (Barry et al., 2008; Caro et al., 2017, Caro et al., 2017), music therapy (Cibrian et al., 2017), university student portal (Fabri et al., 2016; MacLeod, 2010), and communication aids (Checkley et al., 2010; Cheng et al., 2010). These technologies influence every area of life.

Improvements to quality of life are evidenced in reduced anxiety and stress when working with technology, increased options to facilitate interaction with others, and improved tools that aid in pursuing education and employment (Bahiss et al., 2010; Checkley et al., 2010; Fabri et al., 2016; Hatfield et al., 2018). Overall, better usability can result in greater enjoyment during user interactions across this population, as well as improvements to neurotypical responding (Millen et al., 2010).

Telehealth: EHR Systems

It is not difficult to see the value reflected by lives saved and quality of life in relation to health. Errors such as delaying treatment or incorrect medications can be fatal to patients and, therefore, pose a large concern.

As EHR systems become more usable, more people are willing to interact with them. Consequently, remote or isolated users have enjoyed increasing access to healthcare (Czaja et al., 2015).

If improved EHR systems can effectively reduce treatment errors and encourage individuals to seek medical attention when they otherwise may not, incorporating better usability practices into telehealth has the potential to save lives (Mchome et al., 2010).

Cybersecurity: Graphical Authentication

Previous research has shown that 31% of users use the same passwords for all accounts, while 43% of users have never changed their password (Europe, 2008). If systems enforce strong password requirements to protect sensitive account information, usability and security compete. When users elect to use weak passwords, they ultimately compromise the infrastructure of a system. As a result, the security of the system may fail (Cazier & Medlin, 2006; Dawson & Stinebaugh, 2010). Better usability can produce stronger security. This can help to prevent theft, which will reduce financial and emotional stress.

Data Collection and Interpretation Challenges

A reflection on the literature reveals that researchers are concerned about collecting representative data, and simply gaining access to data can be a major challenge. Within the transportation domain, the challenge is to reflect natural driving behavior while keeping participants safe. This typically is achieved by completing studies within simulators. However, even the most realistic driving simulator cannot be considered "a perfect surrogate" for the road (Bédard et al., 2010).

The other domains struggle to access data due to barriers like skepticism, low verbal ability, and the need to maintain privacy. Common obstacles to data collection for aged adults include skepticism of recruitment methods, difficulty staying on track, and reluctance to accept new technology, although acceptance has improved over time (Rubin & Chisnell, 2008).

ASD researchers must consider risk factors like sensory overstimulation and visual fixation in order to prevent unnecessary stress during testing. The need to reduce these risks has led to increased activity in usability research for these individuals. However, it is often difficult to collect usability data due to low verbal abilities or due to anxiety and frustration induced by participatory design (Hirano et al., 2010).

Within EHR systems, a potential barrier to data collection is patient privacy. Similarly, cybersecurity also must navigate privacy concerns carefully. Uniquely, however, the technical needs of cybersecurity often drive design changes beyond usability. Gathering real-world cybersecurity data is difficult, given the close connection between privacy and security.

Clearly, the ability to compare usability findings with a standard or benchmarked data helps with appropriate interpretation. In the cybersecurity domain, benchmarking data and efficiency standards for authentication schemes have been impactful. However, it is often the case that standards are not available to researchers (i.e., in aging and ASD). In telehealth, researchers are dependent upon expert opinion to drive design decisions. Surprisingly, benchmarked data is available in the transportation domain, but it is not being referenced during data interpretation.

Organizations such as the National Highway Traffic Safety Association provide benchmarks for various aspects of in-vehicle information system usability pertaining to efficiency, driver workload, and related concepts. However, such benchmarks are not commonly referred to in current usability testing for in-vehicle information systems literature.

For aging populations, benchmarks or standards for usability were found to be lacking in the literature. Similarly, usability research for autistic individuals was found to lack consistent measures. This inconsistency probably reflects the varying severity of ASD.

A lack of benchmarks or standards emerges in the domain of EHR systems as well. Design decisions are often based on expert opinions rather than on usability findings. Within the graphical authentication cybersecurity literature, the methods employed to determine a system's resistance to cyber-attacks are often inconsistent (Cain et al., 2017). Benchmarking of authentication measures is rare (Cain et al., 2018), but some standards for authentication efficiency do exist (Braz & Robert, 2006).

Usability researchers are motivated to further improve their practices. It is clear that future research will need to discover clever ways to overcome obstacles that prevent data collection and to improve data interpretation (Table 2). Maybe researchers can work to collect more consistent measurements across their research domains. These data can lead to standards or benchmarks that can help researchers produce more impactful interpretations.

CONCLUSION

Usability testing is often employed to find practical design issues. However, usability researchers who publish their findings must be able to situate their design decisions' usability performance compared to others—enabling theory development. Further, they could benefit from benchmarking data to determine whether a product meets a standard threshold of performance. Researchers must start collecting more consistent data across studies.

Usability conventions are emerging within domain-specific areas. We found unique adoption and measurement practices. So, future usability training materials ought to reflect this grounding. Usability research needs to evolve from providing general to more specific guidance within popular application domains (i.e., saving time and limited resources). In addition, usability research offers real value (e.g., from minimizing the amount of required training to saving lives). However, we were surprised by the lack of empirical evidence showing or estimating this value.

TABLE 2: Key Findings: Strengths, Weaknesses, and Future Recommendations

Domain	Strengths	Weaknesses	Future Recommendations
Transportation (in-vehicle information systems)	• Existing benchmarks • Validity of driving simulators	• Complications as a secondary task performance	• Reference benchmarks in data interpretation • Design menus for familiarity
Aging populations	• Extensive empirical articles • Broad in testing, scope, and evaluation	• Lack of benchmarks • Little capitalization on crystallized intelligence	• More comprehensive understanding of acceptance toward new technology • Learnability measurements
Autistic populations	• Wide range of research applications	• Lack of benchmarks • User variability	• Establish criteria/standards to minimize overstimulation and fixation
Telehealth (EHR systems)	• Consistency of issues across EHR systems	• Lack of benchmarks • Reliance on "expert opinions" rather than objective testing	• Implement design decisions that address widely recognized usability issues
Cybersecurity (graphical authentication)	• Extensive measures of login efficiency and accuracy (effectiveness) • Widespread use of surveys to measure satisfaction • Recognition of security weaknesses	• Inconsistent standards for success rate frequency • Not clearly stated when unsuccessful login trial data was excluded from login time analysis • Inconsistent measures of security and password assignment (user- or system-assigned)	• Establish specific protocols for efficiency and effectiveness • Collect more learnability and memorability data

Note. The uniqueness of each domain is exemplified by varying areas of strength and weakness. While there is a significant lack of usability benchmarking standards in general, some domains (i.e., transportation) contain more than others. Along with a lack of benchmarks, inconsistency in measurements—perhaps as both a cause and effect of weak benchmarking—was shown to be a prevailing weakness. EHR = electronic health record.

Usability assessment, as expected, centered on the three classic dimensions of effectiveness, efficiency, and satisfaction. Not surprisingly, as previous literature mentioned, learnability and memorability are still largely unrepresented (Hornbæk, 2006).

The lack of domain-specific usability requirements produced a large amount of variability in how usability dimensions were measured. This makes comparisons across the literature difficult. Within transportation, for example, studies of in-vehicle information systems apply task completion interchangeably as both a measure of efficiency and effectiveness (shown in Table 1).

We were concerned about the lack of domain-specific standards. Without standards of measures and benchmarks, the evidence being presented can easily be biased to support the proposed solution. Lack of uniformity in usability measurements stands in the way of meaningful comparisons—and, ultimately, of better practices. Within graphical authentication research, for example, there is inconsistency in the measurement of login attempts. Some studies measure the number of login attempts until a successful login, while others measure whether a login was successfully completed within a predetermined number of attempts.

Ultimately, with the development of more prototypes within common domains, we need to start translating the successes and failures of our design decisions into theory. If usability researchers can agree on similar constructs, we can use benchmarking data to make meaningful cross-study insights.

KEY POINTS

- Usability is very domain specific, but is often taught and highlighted in a generic fashion.
- Usability measurement varies within research domains, making data interpretation difficult.
- Simply gaining access to data can be a major challenge for usability researchers.
- Usability standards and benchmarking data were often lacking, making data interpretation difficult.

ACKNOWLEDGMENT

All authors but the first and last are listed in alphabetical order and were responsible for conducting separate literature reviews. William Lehman reviewed the ASD population, Wyatt McManus reviewed transportation, Sarah Powers reviewed telehealth, Lauren Tiller reviewed cybersecurity, and James Unverricht reviewed the aging population. Janine Mator wrote the first draft of the manuscript and integrated the separate reviews. Jeremiah Still developed the major reflective themes, wrote some subsections, and completed a major revision.

ORCID iD

Jeremiah D. Still https://orcid.org/0000-0002-3060-4417

REFERENCES

Akatsu, H., & Miki, H. (2004). Usability research for the elderly people. *Oki Technical Review*, *71*, 54–57.

Anderson, G. (2016). Technology trends among mid-life and older Americans. *AARP Research*.

Artoni, S., Buzzi, M. C., Buzzi, M., Ceccarelli, F., Fenili, C., Rapisarda, B., & Tesconi, M. (2011). *Designing ABA-based software for low-functioning autistic children* [Conference session]. International Conference on Advances in New Technologies, Interactive Interfaces, and Communicability, Springer, Berlin, 230–242.

Artoni, S., Buzzi, M. C., Buzzi, M., & Fenili, C. (2010). Didactic software for autistic children. In *International Conference on Advances in New Technologies, Interactive Interfaces, and Communicability* (pp. 73–80): Springer.

Bahiss, K., Cunningham, S. J., & Smith, T. (2010). *Investigating the usability of social networking sites for teenagers with autism* [Conference session]. Proceedings of the 11th International Conference of the NZ Chapter of the ACM Special Interest Group on Human-Computer Interaction, Auckland, New Zealand, 5–8.

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson, R. C., White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L. -C., Harrington, R., Lopez, M., Fitzgerald, R. T., Hewitt, A., . . .Dowling, N. F. (2018). Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR. Surveillance Summaries*, *67*, 1–23.

Barón, A., & Green, P. (2006). Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review (No. UMTRI-2006-5). University of Michigan, Transportation Research Institute.

Barry, M., Kehoe, A., & Pitt, I. (2008). Usability evaluation of educational game software for children with autism. In *EdMedia+ innovate learning* (pp. 1366–1370). Association for the Advancement of Computing in Education (AACE).

Bédard, M. B., Parkkari, M., Weaver, B., Riendeau, J., & Dahlquist, M. (2010). Assessment of driving performance using a simulator protocol: Validity and reproducibility. *American Journal of Occupational Therapy*, *64*, 336–340.

Bevan, N. (2001). International standards for HCI and usability. *International Journal of Human-Computer Studies*, *55*, 533–552.

Bhutta, K. S., & Huq, F. (1999). Benchmarking—best practices: An integrated approach. *Benchmarking: An International Journal*, *6*, 254–268.

Bianchi, A., Oakley, I., & Kim, H. (2016). PassBYOP: Bring your own picture for securing graphical passwords. *IEEE Transactions on Human-Machine Systems*, *46*, 380–389.

Bias, R. G., & Mayhew, D. J. (1994). *Cost-justifying usability*. Morgan Kaufmann.

Booker, J. E., Chewar, C. M., & McCrickard, D. S. (2004). *Usability testing of notification interfaces: Are we focused on the best metrics?* [Conference session]. Proceedings of the 42nd Annual Southeast Regional Conference, ACM, Huntsville, AL, United States, 128–133.

Braz, C., & Robert, J. M. (2006). *Security and usability: The case of the user authentication methods* [Conference session]. Proceedings of the 18th International Conference of the Association Francophone Interaction Homme-Machine, Montreal, QC, Canada, 199–203.

Brosnan, M., Parsons, S., Good, J., & Yuill, N. (2016). How can participatory design inform the design and development of innovative technologies for autistic communities? *Journal of Assistive Technologies*, *10*, 115–120.

Cain, A. A., Edwards, M. E., & Still, J. D. (2018). An exploratory study of cyber hygiene behaviors and knowledge. *Journal of Information Security and Applications*, *42*, 36–45.

Cain, A. A., Werner, S., & Still, J. D. (2017). *Graphical authentication resistance to over-the-shoulder-attacks* [Conference session]. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, ACM, Denver, CO, United States, 2416–2422.

Caro, K., Tentori, M., Martinez-Garcia, A. I., & Alvelais, M. (2017). Using the FroggyBobby exergame to support eye-body coordination development of children with severe autism. *International Journal of Human-Computer Studies*, *105*, 12–27.

Caro, K., Tentori, M., Martinez-Garcia, A. I., & Zavala-Ibarra, I. (2017). FroggyBobby: An exergame to support children with motor problems practicing motor coordination exercises during therapeutic interventions. *Computers in Human Behavior*, *71*, 479–498.

Cazier, J. A., & Medlin, B. D. (2006). Password security: An empirical investigation into e-commerce passwords and their crack times. *Information Systems Security*, *15*, 45–55.

Charness, N., & Boot, W. R. (2009). Aging and information technology use: Potential and barriers. *Current Directions in Psychological Science*, *18*, 253–258.

Checkley, R., Hodge, N., Chantler, S., Reidy, L., & Holmes, K. (2010). What children on the autism spectrum have to 'say' about using high-tech voice output communication aids (VOCAs) in an educational setting. *Journal of Assistive Technologies*, *4*, 25–37.

Cheek, P., Nikpour, L., & Nowlin, H. D. (2005). Aging well with smart technology. *Nursing Administration Quarterly*, *29*, 329–338.

Chen, M. C., Wu, T. F., Lin, Y. L., Tasi, Y. H., & Chen, H. C. (2009). The effect of different representations on reading digital text for students with cognitive disabilities. *British Journal of Educational Technology*, *40*, 764–770.

Cheng, Y., Chiang, H. C., Ye, J., & Cheng, L. H. (2010). Enhancing empathy instruction using a collaborative virtual learning environment for children with autistic spectrum conditions. *Computers & Education*, *55*, 1449–1458.

Chiasson, S., Biddle, R., & Van Oorschot, P. C. (2007). *A second look at the usability of click-based graphical passwords* [Symposium]. Proceedings of the 3rd Symposium on Usable Privacy and Security, ACM, Pittsburgh, PA, United States. 1–12.

Chun, Y. J., & Patterson, P. E. (2012). A usability gap between older adults and younger adults on interface design of an Internet-based telemedicine system. *Work*, *41*, 349–352.

Cibrian, F. L., Peña, O., Ortega, D., & Tentori, M. (2017). BendableSound: An elastic multisensory surface using touch-based interactions to assist children with severe autism during music therapy. *International Journal of Human-Computer Studies*, *107*, 22–37.

Clamann, M., & Kaber, D. B. (2004). Applicability of usability evaluation techniques to aviation systems. *The International Journal of Aviation Psychology*, *14*, 395–420.

Cobb, S., Beardon, L., Eastgate, R., Glover, T., Kerr, S., Neale, H., Parsons, S., Benford, S., Hopkins, E., Mitchell, P., Reynard, G., & Wilson, J. (2002). Applied virtual environments to support learning of social interaction skills in users with Asperger's syndrome. *Digital Creativity*, *13*, 11–22.

Cobb, S. V., Millen, L., Glover, T., Parsons, S., Garib-Penna, S., Weiss, PL., & Giusti, L. (2010). *Integrative approach for designing collaborative technologies for social competence training in children with autism spectrum conditions* [Conference session]. 8th International Conference on Disability, Virtual Reality and Associated Technologies, Valparaíso, Chile, 295–298.

Cockton, G. (2013). Usability evaluation. In D. R. F. Soegaard M (Ed.), *The encyclopedia of human-computer interaction* (2nd ed.). The Interaction Design Foundation. https://www.interactiondesign.org/encyclopedia/usability_evaluation.html.

Czaja, S. J., & Lee, C. C. (2007). The impact of aging on access to technology. *Universal Access in the Information Society*, *5*, 341–349.

Czaja, S. J., Zarcadoolas, C., Vaughon, W. L., Lee, C. C., Rockoff, M. L., & Levy, J. (2015). The usability of electronic personal health record systems for an underserved adult population. *Human Factors*, *57*, 491–506.

Dawson, L. A., & Stinebaugh, J. (2010). *Schemeology for prioritizing cyber-vulnerable critical infrastructure equipment and mitigation strategies* (No. SAND2010-1845). Sandia National Laboratories.

Dumas, J. (2007). The great leap forward: The birth of the usability profession (1988-1993). *Journal of Usability Studies*, *2*, 54–60.

Eklund, J., & Levingston, C. (2008). Usability in Agile development. *UX Research*, 1–7. https://pdfs.semanticscholar.org/349f/3f4af5c9f2499339c132f5cc6d1f56ef3cb9.pdf

Europe, I. (2008). Woman 4 times more likely than men to give passwords for chocolate. Press Release. https://www.infosec.co.uk/page.cfm/T=m/Action=Press/Pre ssID=1071

Fabri, M., Andrews, P. C., & Pukki, H. K. (2016). Using design thinking to engage autistic students in participatory design of an online toolkit to help with transition into higher education. *Journal of Assistive Technologies*, *10*, 102–114.

Feng, F., Liu, Y., & Chen, Y. (2017). A computer-aided usability testing tool for in-vehicle infotainment systems. *Computers & Industrial Engineering*, *109*, 313–324.

Gao, H., Guo, X., Chen, X., Wang, L., & Liu, X. (2008). *Yagp: Yet another graphical password strategy* [Conference session]. Proceedings of the 2008 Annual Computer Security Applications Conference (ACSAC), IEEE, Austin, TX, United States, 121–129.

Garcia-Sanjuan, F., Jaen, J., & Nacher, V. (2017). Tangibot: A tangible-mediated robot to support cognitive games for ageing people—A usability study. *Pervasive and Mobile Computing*, *34*, 91–105.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident Analysis and Prevention*, *34*, 589–600.

Goldberg, L., Lide, B., Lowry, S., Massett, H. A., O'Connell, T., Preece, J., & Shneiderman, B. (2011). Usability and accessibility in consumer health informatics: Current trends and future challenges. *American Journal of Preventive Medicine*, *40*, S187–S197.

Grassie, P. A., Garcia, M. E., & Fenton, J. L. (2017). *Digital identity guidelines* (Technical Report). NIST Special Publication 800-63-3. 800-63-3.

Grawemeyer, B., & Johnson, H. (2011). Using and managing multiple passwords: A week to a view. *Interacting with Computers*, *23*, 256–267.

Green, P. (1999). *Estimating compliance with the 15-second rule for driver-interface usability and safety* [Conference session]. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 43(18), Sage CA; Los Angeles, CA, SAGE Publications, 987–991.

Greenberg, S., & Buxton, B. (2008). *Usability evaluation considered harmful (some of the time)* [Conference session]. Proceedings of the SIGCHI Conference on Human factors in Computing Systems, New York, NY, ACM Press, 111–120.

Grudin, J. (2008). A moving target: The evolution of HCI. In A. Sears & J. A. Jacko (Eds.), *The human computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (2nd ed.,pp. 1–24). Lawrence Erlbaum Associates.

Grynszpan, O., Martin, J. C., & Nadel, J. (2008). Multimedia interfaces for users with high functioning autism: An empirical investigation. *International Journal of Human-Computer Studies*, *66*, 628–639.

Hanson, V. L., & Crayne, S. (2005). Personalization of web browsing: Adaptations to meet the needs of older adults. *Universal Access in the Information Society*, *4*, 46–58.

Harvey, C. (2009). *Development of a method for evaluating the usability of in-vehicle information systems (IVISs)* [Conference session]. IFIP Conference on Human-Computer Interaction, Uppsala, Sweden, 856–859.

Harvey, C., Stanton, N. A., Pickering, C. A., McDonald, M., & Zheng, P. (2011). A usability evaluation toolkit for in-vehicle information systems (IVISs). *Applied Ergonomics*, *42*, 563–574.

Haslwanter, J. D., Fitzpatrick, G., & Miesenberger, K. (2018). Key factors in the engineering process for systems for aging in place contributing to low usability and success. *Journal of Enabling Technologies*, *12*, 186–196.

Hatfield, M., Falkmer, M., Falkmer, T., & Ciccarelli, M. (2018). Process evaluation of the BOOST-A™ transition planning program for adolescents on the autism spectrum: A strengths-based approach. *Journal of Autism and Developmental Disorders*, *48*, 377–388.

Hayashi, E., Dhamija, R., Christin, N., & Perrig, A. (2008). *Use your illusion: Secure authentication usable anywhere* [Symposium]. Proceedings of the 4th Symposium on Usable Privacy and Security, Pittsburgh, PA, United States, 35–45.

Herley, C., & Van Oorschot, P. (2011). A research agenda acknowledging the persistence of passwords. *IEEE Security & Privacy*, *10*, 28–36.

Hernández-Encuentra, E., Pousada, M., & Gómez-Zúñiga, B. (2009). ICT and older people: Beyond usability. *Educational Gerontology*, *35*, 226–245.

Hirano, S. H., Yeganyan, M. T., Marcu, G., Nguyen, D. H., Boyd, L. A., & Hayes, G. R. (2010). *vSked: Evaluation of a system*

to support classroom activities for children with autism [Conference session]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Atlanta, GA, United States. 1633–1642.
Hitchcock, D. R., Lockyer, S., Cook, S., & Quigley, C. (2001). Third age usability and safety—An ergonomics contribution to design. *International Journal of Human-Computer Studies*, *55*, 635–643.
Hoonakker, P., Bornoe, N., & Carayon, P. (2009). Password authentication from a human factors perspective: Results of a survey among end-users. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53, 459–463.
Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, *64*, 79–102.
Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*, 611–624.
Hsieh, K. L., Fanning, J. T., Rogers, W. A., Wood, T. A., & Sosnoff, J. J. (2018). A fall risk mhealth app for older adults: Development and usability study. *JMIR Aging*, *1*, e11569.
Inglesant, P. G., & Sasse, M. A. (2010). *The true cost of unusable password policies: Password use in the wild*, ACM.
International Electrotechnical Committee (IEC). (2015). IEC 62366-1:2015: Medical Devices—Part 1: Application of Usability Engineering to Medical Devices, Geneva, Switzerland.
Institute of Medicine (U. S.). (2012). *Health IT and patient safety: Building safer systems for better care*. National Academies Press.
International Organization for Standardization. (1998). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs): Part 11: Guidance on usability*.
ISO 62366. (2008). ISO 62366: 2008 Medical Devices–Application of Usability Engineering to Medical Devices.
Jacko, J. A., Barreto, A. B., Marmet, G. J., Chu, J. Y., Bautsch, H. S., Scott, I. U., & Rosa,R. H. (2000). *Low vision: the role of visual acuity in the efficiency of cursor movement* [Conference session]. Proceedings of the Fourth International ACM Conference on Assistive Technologies, ACM, Arlington, VA, United States. 1–8.
Jeon, M., Zhang, R., Lehman, W., Fakhrhosseini, S., Barnes, J., & Park, C. H. (2015). *Development and evaluation of emotional robots for children with autism spectrum disorders* [Conference session]. International Conference on Human-Computer Interaction, Los Angeles, CA, Springer International, 372–376.
Johnson, J. (2008). *GUI bloopers 2.0: Common user interface design don'ts and dos*. Elsevier.
Jordan, P. W., Thomas, B., McClelland, I. L., & Weerdmeester, B. (Eds.). (2014). *Usability evaluation in industry*. CRC Press.
Keith, S. (2010). Designing with older car drivers: Seeking out aspirations and needs. *Quality in Ageing and Older Adults*, *11*, 38–47.
Khan, S., Tahir, M. N., & Raza, A. (2013). *Usability issues for smartphone users with special needs—Autism* [Conference session]. 2013 International Conference on Open Source Systems and Technologies, IEEE, Lahore, Pakistan, 107–113.
Khot, R. A., Kumaraguru, P., & Srinathan, K. (2012). *WYSWYE: Shoulder surfing defense for recognition based graphical passwords* [Conference session]. Proceedings of the 24th Australian Computer-Human Interaction Conference, Melbourne, Australia, 285–294.
Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (Technical Report No. DOT HS 810 594). National Highway Traffic Safety Administration. https://vtechworks.lib.vt.edu/bitstream/handle/10919/55090/DriverInattention.pdf?sequence=1%26isAllowed=y.
Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research methods in human-computer interaction*. John Wiley & Sons.
Lee, H. C., Cameron, D., & Lee, A. H. (2004). Assessing the driving performance of older adult drivers: On-road versus simulated driving. *Accident Analysis & Prevention*, *35*, 797–803.
Leu, E. (2017). Authentication Trends for 2017. Retrieved March 11, 2019, from https://www.upwork.com/hiring/for-clients/authentication-trends/.
Lewis, J. R. (1992). Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the Human Factors and Ergonomics Society 36th Annual Meeting* (Vol. 36, pp. 1259–1263). Human Factors Society.
Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*, 57–78.
Li, R., Chen, Y. V., Sha, C., & Lu, Z. (2017). Effects of interface layout on the usability of in-vehicle information systems and driving safety. *Displays*, *49*, 124–132.
Lund, A. (2011). *User experience management: Essential skills for leading effective UX teams*. Morgan Kaufmann.
Ma, J., & Du, Y. (2017). *Study on the evaluation method of in-vehicle gesture control* [Conference session]. 2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE), IEEE, 145–148.
MacDonald, C. M., & Atwood, M. E. (2013). Changing perspectives on evaluation in HCI: Past, present, and future. In *CHI '13 extended abstracts on human factors in computing systems* (pp. 1969–1978). ACM.
MacLeod, A. (2010). 'Welcome to my first rant!' Report on a participatory pilot project to develop the 'AS portal', an online peer support network for higher education students on the autism spectrum. *Journal of Assistive Technologies*, *4*, 14–24.
Matthews, G., Davies, D. R., Stammers, R. B., & Westerman, S. J. (2000). *Human performance: Cognition, stress, and individual differences*. Psychology Press.
Mayhew, D. (2008). User-experience design: The evolution of a multi-disciplinary approach. *Journal of Usability*, *3*, 99–102.
McCarthy, S., Sayers, H., & McKevitt, P. (2007). *Investigating the usability of PDAs with ageing users* [Conference session]. Proceedings of the 21st British HCI Group Annual Conference on People and Computers, 2, BCS Learning & Development Ltd, Lancaster, England, 67–70.
Mchome, S., Sachdeva, S., & Bhalla, S. (2010). A brief survey: Usability in healthcare. IEEE 2010 International Conference on Electronics and Information Engineering, 1, V1–463.
Melander-Wikman, A., Fältholm, Y., & Gard, G. (2007). Safety vs. privacy: Elderly persons' experiences of a mobile safety alarm. *Health and Social Care in the Community*, *16*, 337–346.
Millen, L., Edlin-White, R., & Cobb, S. (2010). *The development of educational collaborative virtual environments for children with autism* [Conference session]. Proceedings of the 5th Cambridge Workshop on Universal Access and Assistive Technology, Cambridge, 7–13.
Miriovsky, B. J., Shulman, L. N., & Abernethy, A. P. (2012). Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. *Journal of Clinical Oncology*, *30*, 4243–4248.
Mitsopoulos-Rubens, E., Trotter, M. J., & Lenné, M. G. (2011). Usability evaluation as part of iterative design of an in-vehicle information system. *IET Intelligent Transport Systems*, *5*, 112–119.
Monibi, M., & Hayes, G. R. (2008). *Mocotos: mobile communications tools for children with special needs* [Conference session]. Proceedings of the 7th International Conference on Interaction Design and Children, ACM, Chicago, IL, United States, 121–124.
National Center for Statistics and Analysis. (2017). *Distracted Driving 2015* (Traffic Safety Facts Research Note. Report No. DOT HS 812 381). National Highway Traffic Safety Administration.
Nielsen, J. (1995). 10 usability heuristics for user interface design. *Nielsen Norman Group*, *1*.
Nilsson, M., Johansson, S., & Håkansson, M. (2003). *Nostalgia: An evocative tangible interface for elderly users*. CHI'03 Extended Abstracts on Human Factors in Computing Systems, ACM, 964–965.
Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. Basic Books.
O'Connell, R. T., Cho, C., Shah, N., Brown, K., & Shiffman, R. N. (2004). Take note (s): differential EHR satisfaction with two implementations under one roof. *Journal of the American Medical Informatics Association*, *11*, 43–49.

Oksman, V. (2006). Young people and seniors in Finnish “mobile information society”. *Journal of Interactive Media in Education*, *2*.

Otjacques, B., Krier, M., Feltz, F., Ferring, D., & Hoffmann, M. (2009). *Helping older people to manage their social activities at the retirement home* [Conference session]. Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology, British Computer Society, Cambridge, England, 375–380.

Paivio, A. (2013). *Imagery and verbal processes*. Psychology Press.

Pak, R., Czaja, S. J., Sharit, J., Rogers, W. A., & Fisk, A. D. (2008). The role of spatial abilities and age in performance in an auditory computer navigation task. *Computers in Human Behavior*, *24*, 3045–3051.

Peng, Y., Boyle, L. N., & Hallmark, S. L. (2013). Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. *Accident Analysis & Prevention*, *50*, 628–634.

Petrovčič, A., Taipale, S., Rogelj, A., & Dolničar, V. (2018). Design of mobile phones for older adults: An empirical analysis of design guidelines and checklists for feature phones and smartphones. *International Journal of Human–Computer Interaction*, *34*, 251–264.

Politis, Y., Olivia, L., Olivia, T., & Sung, C. (2017). Involving people with autism in development of virtual world for provision of skills training. *International Journal of E-Learning & Distance Education*, *32*, 1–16.

Politis, Y., Sung, C., Goodman, L., & Leahy, M. (2019). Conversation skills training for people with autism through virtual reality: Using responsible research and innovation approach. *Advances in Autism*, *1*.

Ranney, T. A., Baldwin, G. H., Parmer, E., Domeyer, J., Martin, J., & Mazzae, E. N. (2011). *Developing a test to measure distraction potential of in-vehicle information system tasks in production vehicles* (No. HS-811 463). National Highway Traffic Safety Administration.

Reed, M. P., & Green, P. A. (1999). Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task. *Ergonomics*, *42*, 1015–1037.

Rojas, C. L., & Seckman, C. A. (2014). The informatics nurse specialist role in electronic health record usability evaluation. *CIN: Computers, Informatics*. *Nursing*, *32*, 214–220.

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: How to plan, design and conduct effective tests*. John Wiley & Sons.

Sarkar, U., Gourley, G. I., Lyles, C. R., Tieu, L., Clarity, C., Newmark, L., Singh, K., & Bates, D. W. (2016). Usability of commercially available mobile applications for diverse patients. *Journal of General Internal Medicine*, *31*, 1417–1426.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, *49*, 304–313.

Schumacher, R. M., Berkowitz, L., Abramson, P., & Liebovitz, D. (2010, September). *Electronic health records: Physician’s perspective on usability* [Conference session]. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54, 816-820, Sage CA, Los Angeles, CA, SAGE Publications.

Segall, N., Saville, J. G., L’Engle, P., Carlson, B., Wright, M. C., Schulman, K., & Tcheng, J. E. (2011). *Usability evaluation of a personal health record* [Symposium]. AMIA Annual Symposium Proceedings American Medical Informatics Association, Washington, D.C., United States.

Shneiderman, B. (2011). Tragic errors: Usability and electronic health records. *Interactions*, *18*, 60–63.

Singh, J., Lutteroth, C., & Wünsche, B. C. (2010, July). *Taxonomy of usability requirements for home telehealth systems* [Conference session]. Proceedings of the 11th International Conference of the NZ Chapter of the ACM Special Interest Group on Human-Computer Interaction, ACM, Auckland, New Zealand, 29–32.

Sixsmith, A. (2013). *Technology and the challenge of aging*. Technologies for Active Aging, Springer, 7–25.

Smelcer, J. B., Miller-Jacobs, H., & Kantrovich, L. (2009). Usability of electronic medical records. *Journal of Usability Studies*, *4*, 70–84.

Society of Automotive Engineers. (1998). SAE Recommended Practice for Calculating the Time to Complete In-Vehicle Navigation and Route Guidance Tasks (SAE J2365), Committee Draft of November 23. Society of Automotive Engineers

Society of Automotive Engineers. (1999). SAE Recommended Practice for Navigation and Route Guidance Function Accessibility While Driving (SAE J2364), Committee Draft of June 8. Society of Automotive Engineers.

Soegaard, M. (2012). The history of usability: From simplicity to complexity. *Smashing Magazine*. Retrieved April 19, 2019, from http://www.smashingmagazine.com/2012/05/23/the-history-of-usability-from-simplicity-to-complexity.

Sonderegger, A., Schmutz, S., & Sauer, J. (2016). The influence of age in usability testing. *Applied Ergonomics*, *52*, 291–300.

Stevens, A., Quimby, A., Board, A., Kersloot, T., & Burns, P. (2002). *Design guidelines for safety of in-vehicle information systems*. TRL Limited.

Still, J. D., Cain, A. A., & Schuster, D. (2017). Human-centered authentication guidelines. *Information & Computer Security*, *25*, 437–453.

Stuster, J. (2006). *The human factors and ergonomics society: Stories from the first 50 years*. HFES.

Sumie, M., Li, C., & Green, P. (1998). *Usability of menu-based interfaces for motor vehicle secondary functions.*(Technical Report UMTRI-97-19). University of Michigan Transportation Research Institute.

Sun, H. M., Chen, Y. H., Fang, C. C., & Chang, S. Y. (2012). *PassMap: a map based graphical-password authentication system* [Conference session]. Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, ACM, Seoul, Republic of Korea, 99–100.

Tiller, L. N., Angelini, C. A., Leibner, S. C., & Still, J. D. (2019). Explore-a-nation: Combining graphical and alphanumeric authentication. In A. Moallem (Ed.), *HCI for cybersecurity, privacy and trust, HCII* (Vol. 11594, pp. 81–95). Springer.

Van Laarhoven, T., Carreon, A., Bonneau, W., & Lagerhausen, A. (2018). Comparing mobile technologies for teaching vocational skills to individuals with autism spectrum disorders and/or intellectual disabilities using universally-designed prompting systems. *Journal of Autism and Developmental Disorders*, *48*, 2516–2529.

Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R., & Zhao, Q. (2015). Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron*, *88*, 604–616.

Wang, L., Rau, P. L. P., & Salvendy, G. (2011). Older adults' acceptance of information technology. *Educational Gerontology*, *37*, 1081–1099.

Weinberg, G., Harsham, B., & Medenica, Z. (2011). *Evaluating the usability of a head-up display for selection from choice lists in cars* [Conference session]. Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ACM, 39–46.

Wiedenbeck, S., Waters, J., Birget, J. C., Brodskiy, A., & Memon, N. (2005). PassPoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, *63*, 102–127.

Wiedenbeck, S., Waters, J., Sobrado, L., & Birget, J. C. (2006). *Design and evaluation of a shoulder-surfing resistant graphical password scheme* [Conference session]. Proceedings of the Working Conference on Advanced Visual Interfaces, Venice, Italy, 177–184.

Wixon, D. (2003). Evaluating usability methods: Why the current literature fails the practitioner. *Interactions*, *10*, 28–34.

Yan, J., Blackwell, A., Anderson, R., & Grant, A. (2004). Password memorability and security: Empirical results. *IEEE Security & Privacy Magazine*, *2*, 25–31.

Zakari, H. M., Ma, M., & Simmons, D. (2014). *A review of serious games for children with autism spectrum disorders (ASD)* [Conference session]. International Conference on Serious Games Development and Applications, Springer, Cham, 93–106.

Zangooei, T., Mansoori, M., & Welch, I. (2012). *A hybrid recognition and recall based approach in graphical passwords* [Conference session]. Proceedings of the 24th Australian Computer-Human Interaction Conference, ACM, 665–673.

Ziefle, M., & Bay, S. (2005). How older adults meet complexity: aging effects on the usability of different mobile phones. *Behaviour & Information Technology*, *24*, 375–389.

Janine D. Mator is a graduate student at Old Dominion University. Her research explores usability research and testing. She is also interested in research at the intersection of human–computer interaction and cognition.

William E. Lehman is a graduate student at Old Dominion University. His research focuses on human–automation interaction. His past research has included the development of a social robot to train emotion recognition for children on the autism spectrum; the influence of expertise on mental rotation; and usability testing for in-vehicle touch screen systems.

Wyatt McManus is a graduate student at Old Dominion University. His research examines how humans interact with autonomous and manually controlled vehicles in emergency situations. Specifically, he is interested in the way that humans interact with autonomous vehicle systems in response to cyber-attacks.

Sarah Powers is a graduate student at Old Dominion University. Her research focuses on the effects of interruptions on human performance, with research interests in the healthcare domain.

Lauren Tiller is a graduate student at Old Dominion University. Her research focuses on the human aspects of cybersecurity. Her past research has introduced a new graphical authentication scheme and explored the trade-off between usability and security.

James R. Unverricht is a graduate student at Old Dominion University. His research interests include attention, calibration, and human–automation interaction within the surface transportation domain. He is also interested in how special populations, such as young or aged drivers, interact with various in-vehicle technologies.

Jeremiah D. Still is an assistant professor at Old Dominion University. His Psychology of Design laboratory explores the relationship between human cognition and technology; specifically, he is focusing on usable cybersecurity, visual attention, and intuitive design.

Date received: May 3, 2019
Date accepted: November 12, 2019