Investigating Best Practices for Remote Summative Usability Testing with People with Mild to Moderate Dementia

RACHEL WOOD, EMMA DIXON, SALMA ELSAYED-ALI, EKTA SHOKEEN, AMANDA LAZAR, and JONATHAN LAZAR, Information Studies, University of Maryland, College Park, MD, USA

People with dementia may miss out on the benefits of using technology, because they often find it difficult to use. Usability testing is one method to identify barriers and areas for improvement in technology. Unfortunately, usability testing is often not conducted with people with dementia, independent of their caregivers. Difficulty recruiting local participants with dementia who regularly use technology further compounds the problem. Remote methods have been proposed as one approach to recruiting hard-to-reach populations. Currently, it is unclear how to effectively conduct remote summative usability testing with people with dementia. We recruited 15 participants. Five took part in the pilot study and 10 participated in the main study. We identify best practices and make suggestions for remote summative usability tests with people who have mild to moderate dementia, independent of caregivers. We discuss our findings in three sections: (1) logistics for planning remote summative usability testing, (2) approaches for conducting remote summative usability testing, including modifications of research methods, and (3) considerations when evaluating findings from remote summative usability sessions. We also present modified usability testing methods we developed to meet the unique needs of users with mild to moderate dementia, and summarize lessons learned and new directions for research on this topic.

CCS Concepts: • **Human-centered computing** → *HCI design and evaluation methods*; *Accessibility*; *Interaction design*; *Interaction design process and methods*;

Additional Key Words and Phrases: Remote usability, remote summative, UX, dementia, disability, remote methods

ACM Reference format:
Rachel Wood, Emma Dixon, Salma Elsayed-Ali, Ekta Shokeen, Amanda Lazar, and Jonathan Lazar. 2021. Investigating Best Practices for Remote Summative Usability Testing with People with Mild to Moderate Dementia. *ACM Trans. Access. Comput.* 14, 3, Article 11 (July 2021), 26 pages.
https://doi.org/10.1145/3460942

This work was funded in part by the National Institute on Disability, Independent Living and Rehabilitation Research (NIDILRR) at the Administration for Community Living, U.S. Department of Health and Human Services under Grant no. 90REGE0008 and by the U.S. Department of Education Grant no. H421A150006 from and by the National Science Foundation Graduate Research Fellowship Program under Grant no. DGE 1840340. The opinions herein are those of the authors and not necessarily those of the funding agencies.

Authors' addresses: R. Wood, E. Dixon, S. Elsayed-Ali, E. Shokeen, A. Lazar, and J. Lazar, College of Information Studies, Trace Research and Development Center, HCIL, University of Maryland, 4130 Campus Dr 4th floor, College Park, MD 20742; emails: {rewood, eedixon, sea, eshokeen, lazar, jlazar}@umd.edu.

1936-7228/2021/07-ART11 $15.00
https://doi.org/10.1145/3460942

Corrected Version of Record. V.1.1. Published July 30, 2021.

1 INTRODUCTION

As technology increasingly becomes a part of everyday life, current research indicates poor user adoption among people with dementia due to confusion about how to use it [36], difficulty understanding the potential benefit of using a new device [32, 33], and low computer literacy [31]. Usability issues for people with dementia have emerged as a significant barrier, including issues with touch screens [67], device size [22], error recovery [85], user control [85], providing adequate information [85], and lack of personalization [5, 65]. While some researchers may use indirect evaluation methods, such as only involving caregivers [41, 76, 79] or the use of personas [62, 73, 83], to include derivative perspectives of people with dementia on their behalf into the design, usability testing allows for direct involvement of people with dementia during the development of technologies intended to be used by them. Usability testing is an established technique in user-centered design to improve the ease of use of interfaces for representative users by detecting interface flaws or areas that need improvement [48]. However, many conventional user-centered design practices, including usability testing, often do not include people with dementia. Even fewer include people with dementia independent of caregivers. There is a gap in the current research describing how to adapt design methods to individuals experiencing ongoing changes in their cognition and communication [14, 36], which underlines the importance of understanding how to apply existing usability testing techniques to the population of people with dementia [11]. The need to develop remote testing methods was further emphasized by the COVID pandemic. This article focuses on identifying challenges and considerations when conducting remote summative usability testing with people in the early stages of dementia.

Usability testing involving people with dementia is challenging for three reasons. First, there is limited literature related to usability testing with this population as independent participants [82]. Most usability tests involving people with dementia include caregivers working alongside them, often helping them to complete tasks. This practice may be more appropriate for those with late-stage dementia. However, it is not representative and is problematic for people in the early stages of dementia, who often interact with technology independent of caregivers. Second, of the published usability studies that included people with dementia, their involvement primarily occurred during the formative stage of usability testing. Few studies involve people with dementia during the summative stage of usability testing, resulting in a lack of understanding on how to conduct it [6, 34, 74, 82]. Third, there is a general lack of understanding of how to leverage the benefits of remote usability testing with people who have dementia, despite the acceptance of these remote methods as a way to increase researchers' ability to work with populations that are difficult to recruit [70] or are stigmatized [52].

The goal of our research is to understand the necessary adaptations and best practices for remote summative usability testing methods involving people who have mild to moderate dementia independent of their caregivers. The three core questions guiding our research are as follows:

(1) What logistics need to be considered when planning remote summative usability testing involving people who have mild to moderate dementia, independent of caregivers?
(2) When conducting remote summative usability testing, what modifications to methods are necessary to better facilitate inclusion of people with mild to moderate dementia, independent of caregivers?
(3) When evaluating findings from a remote summative usability test, what metrics are meaningful for an accurate understanding of the technology needs and preferences of people with mild to moderate dementia, independent of caregivers?

We conducted remote summative usability testing with people with mild to moderate dementia (early-stage), who had prior experience with computers, independent from caregivers. To the best of our knowledge, this is the first study evaluating the challenges and best practices of conducting remote summative usability testing with people with mild to moderate dementia.

2 BACKGROUND

The following sections summarize past work investigating the efficacy of remote summative usability testing methods and an overview of the software that the participants evaluated.

2.1 Remote Usability Testing Methods

Finding enough representative users who meet the inclusion criteria (e.g., a specific disability, technical skill level) is one of the biggest challenges when working with people who have disabilities in either usability testing or other types of research [23, 48]. Remote methods are one way for researchers to overcome this obstacle by recruiting and conducting study activities online [10, 70]. Because researchers are not limited to the same locations as participants, travel-related expenses may be reduced [35] and travel-related logistics avoided. Remote usability testing provides access to international participants from various time zones [26, 35] and can be synchronous [19] or asynchronous [17]. Petrie et al. reported that there is great potential benefit of using remote synchronous usability testing methods to facilitate observation of naturalistic interactions between people with disabilities and their technologies [70]. Furthermore, remote methods provide access to stigmatized populations who may be unwilling to meet and disclose their identity in person [52].

Remote studies also provide participants the ability to complete studies in their natural work or home environment as opposed to an academic or laboratory environment [35, 70]. Furthermore, conducting usability studies in more naturalistic settings is ideal for users with disabilities as it has the added benefit of making them more comfortable using familiar technology in an equally familiar setting [48]. Lazar et al. noted that usability testing using a participant's personal technology in this context is the "easiest and most natural form of usability testing" for the user, and the most difficult for the moderator. As such, remote usability testing using flexible methods may prove beneficial when working with populations, settings, and technologies that can be variable, as well as can support user comfort by conducting the study in lower-stress environments (i.e., home) [48].

Many researchers have examined the feasibility and benefits of remote usability testing methods [4, 26, 35, 70, 86]. Using remote methods with blind participants [70], deaf participants [78], and participants who have aphasia [75] all had promising results. Remote methods have not previously been explored for people who have dementia. One exception is a study on remote reminiscence, which found that people with moderate dementia engage in reminiscence via video conferencing the same as they do in person [43]. When it comes to technology design, researchers have only recently included people with dementia in remote interviews (e.g., [45, 54]). There is an opportunity to leverage the benefits of remote methods for summative usability testing, involving participants with dementia.

2.2 Auto-Personalization and Morphic

The variability between and within people with dementia (i.e., fluctuations hour-to-hour and a longer-term progression of the condition) often prevents a "one-fits-all" approach [55, 77] and even challenges "one-size-fits one" for technology. Research with people who have other cognitive disabilities suggests that adjustable devices [55, 69], personalization, or individually configured technology may also benefit people with dementia [49]. One potential solution for addressing variability and improving accessibility, usability, and adoption of technologies is through the implementation of auto-personalization software. Auto-personalization allows users to save the accessibility and ease of use settings that best fit their needs and easily modify them as needed. Those preferences are then automatically applied to a new interface or other devices [88]. Auto-personalization could help people with dementia by making digital environments easier to navigate and more familiar.

In this remote summative usability study, participants compared the discoverability of accessibility and ease of use features built into Windows to Morphic, one of the most robust examples of auto-personalization software for Windows. The Morphic QuickStrip is a series of buttons located in the bottom right corner of the screen for adjusting accessibility settings. Morphic also presented the opportunity to conduct usability testing on complex software, rather than a simple website. We studied representative user experiences during software installation, computer personalization, and complicated interactions with an unfamiliar technology. A separate, future publication will discuss the usability testing of Morphic and the findings related to specific interface enhancements for the application.

3 RELATED WORK ON USERS WITH DEMENTIA

We separate the literature review on users with dementia into two sections: (1) the necessity of usability testing involving people with dementia and (2) a summary of the usability studies conducted thus far.

3.1 The Necessity of Usability Testing Involving People with Dementia

Research identifies several challenges for people with dementia in their everyday use of new and familiar technologies [53, 66]. These challenges and usability issues stem from complexity [21], learnability [21], reliability [55], convenience [55], and difficulty entering information [55], and all of these challenges impact technology adoption. Users with dementia were unwilling to adopt technology when they perceived it as useless, irrelevant, or unbeneficial [21]. Specific interface aspects that people with dementia struggle using include the size of the screen [22], touch screens [67], limited user control [85], difficulty understanding or recovering from errors [85], nomenclature [74], too many options [18], not enough information [85], and the lack of personalization [5, 65]. Researchers argue that people with dementia need interfaces that support the skills they have, minimize errors and warnings, facilitate feeling independent, and use familiar objects or activities [85]. It also highlights the need for technology to meet the specific needs of people with dementia. Usability testing is essential to understanding and designing for their unique technology requirements.

3.2 Usability Testing Involving People with Dementia

In general, formative usability testing typically happens earlier in the development process, is qualitative, and focuses on general design concepts. Summative usability studies typically evaluate functional software versus formative usability studies which typically involve paper prototypes, mock-ups, or wireframes. Because it focuses on functional software or very high-fidelity prototypes, summative usability testing happens later in development and is focused on evaluating specific design features or interface widgets. Summative usability testing is often interpreted to mean simply measuring task and time performance. While those are typically the metrics used (and may often be viewed that way due to those metrics being identified by the International Organization for Standardization) [89], the real crux of summative usability testing is that it (1) evaluates functional software, rather than the mock-ups or wireframes often used in formative usability testing, and (2) involves representative users attempting representative tasks. While task and time performance metrics (or other quantitative metrics) are frequently used in summative usability studies, the use of such metrics is not a fixed requirement for a summative usability study [48].

To understand the scope of the current body of literature and to identify the gaps in usability testing involving people with dementia, we performed a scoping literature review [71]. Scoping reviews are appropriate when the "purpose of the review is to identify knowledge gaps, scope a body of literature, clarify concepts or to investigate research conduct" [60]. We included only those studies which directly involved people with dementia with no criteria for participants at any specific stage of dementia. This included studies conducted using dyads of caregivers and people with dementia. Additionally, studies had to focus on technology development, either testing existing technologies, prototypes, or designing new technologies. Studies conducted as experimental design, simulation software, or with proxy users, as well as literature reviews were all excluded from this literature review. Sixty-five papers were selected which described the design and evaluation of technologies by people with dementia. Of these papers, we found that the majority of studies (40 papers) used formative methods, such as interviews [27, 59], focus groups [76, 79], storyboarding [54, 56], and workshops [38, 45]. Seventeen papers described summative usability methods. Only eight used both formative and summative methods. No studies reported remote summative usability testing methods with people at any stage of dementia.

Of the 25 papers (the 17 plus 8 papers described above) using summative usability testing, only 9 involved people with dementia independent of caregivers. The remaining 16 summative studies used dyads (i.e., people with dementia partnered with their caregivers). Summative usability studies involving people with dementia (independent of caregivers) may also be rare because of the perceived difficulty obtaining reliable feedback from them [11]. For this reason, usability testing often includes dyads to compensate for a perceived low-technology literacy [51, 55, 56]. Remote usability methods also potentially facilitate involving people with dementia independently, overcoming both recruitment and logistical issues.

Only 2 of the 25 studies described best practices when conducting in-person summative usability testing with people with dementia [11, 31]. These two studies report contradictory results. Gibson et al. evaluated the efficacy of standard usability testing methods with dyads [31]. They found that think-aloud protocols required proactive engagement from the facilitator or caregiver to support the person with dementia's memory while completing tasks [31]. They also found task completion rate and task completion time to be reliable measures [31]. In contrast, Boger et al. found task completion time to be an invalid measure of usability. They attributed this to people with dementia's tendency to take longer on tasks regardless of difficulty and to spend more time on tasks they enjoyed [11]. The two conflicting studies demonstrate the clear need for research into the efficacy of summative usability testing with the population. Building off this work, our study's contribution lies in incorporating remote summative usability testing methods with people who have mild to moderate dementia independent of caregivers, which no previously published research has investigated.

4 RESEARCH METHODOLOGY

To fine-tune our methodology before our main study, we conducted a pilot study with one remote and four in-person participants. Shortly after our pilot, our university prohibited in-person human subjects research in response to the COVID-19 pandemic. At that point, we pivoted to our main usability study, conducted entirely remotely with 10 people who have dementia. We used a typical remote usability method and developed two novel methods, the Remote Access method and Modified Think-Aloud, when technical constraints necessitated more flexible methods, which we detail below.

Table 1. Participant Demographic Information

PID#	AGE	GENDER	USABILITY TEST METHOD	TYPE OF DEMENTIA
PILOT1	61	Female	In Person	Vascular Micro-angiopathy
PILOT2	58	Female	In Person	Alzheimer's
PILOT3	67	Male	Remote: Installation	Vascular
PILOT4	84	Female	In Person	Unknown
PILOT5	63	Female	In Person	Primary Progressive Aphasia
P1	61	Male	Remote: Installation	Alzheimer's
P2	60	Female	Remote: Installation	Vascular
P3	59	Male	Remote: Installation	Vascular and Alzheimer's
P4	58	Male	Remote: Installation	Alzheimer's
P5	73	Female	Remote Access	Vascular
P6	64	Male	Remote: Modified Think-Aloud	Vascular and Alzheimer's
P7	55	Female	Remote Access	Alzheimer's
P8	75	Male	Remote: Modified Think-Aloud	Unknown
P9	73	Male	Remote: Modified Think-Aloud	Alzheimer's
P10	71	Male	Remote: Modified Think-Aloud	Alzheimer's

Pilot# refers to a pilot study participant. P# refers to a main study participant.

4.1 Participants

There were two inclusion criteria for our participants. They needed to have a medical diagnosis of any type of early-stage dementia (e.g., mild to moderate or if they declined to stage themselves, equivalent to those stages in terms of functioning). They also had to use technology in their daily lives. Each of these were self-reported by participants. There are tradeoffs to using self-reported disability status rather than to pursue formal staging of dementia severity by a clinician. Using instruments such as the Mini-Mental State Exam or Montreal Cognitive Assessment can be useful to detect and track changes in cognition. These tools ask questions which assess cognitive functioning in domains such as word recall, attention and calculation, and language abilities. For our purposes—investigating opportunities for increasing technology usefulness, usability, and accessibility for people with dementia—functional capacity can be a more relevant measure as it indicates people's ability to carry out daily activities and tasks. Therefore, we asked participants to self-report based on the categories we drew from National Institutes of Health function-based stage classification [61]. This approach is in alignment with accepted approaches in **human-computer interaction** (**HCI**) which recognize that "It may be hard to get detailed information about the medical status of someone with a disability… because of the sensitivity involved in sharing personal health data" [48].

The average age of participants was 64. For research with participants who have disabilities, it is considered acceptable to have 5–10 participants, as long as all of them meet the specific inclusion criteria [48]. In total, our study consists of 15 participants (5 in the pilot study and 10 in the main study). Table 1 provides the study participants' demographic information.

Participants were recruited from online communities (e.g., large dementia organizations) using word-of-mouth and snow-ball sampling. During the screening process, participants were asked if they had access to a Windows 10 computer that they could install software onto that also had Microsoft Word installed on it, as these were the system requirements to ensure they had a compatible machine. Participants received an electronic copy of the consent form during scheduling. This gave participants additional time to review the consent form and ask any questions about the consent process or study procedures in advance. Participants signed an online version of the consent form using Qualtrics before participation in the study. All participants directly provided their contact information to be included in the study, demonstrating the capacity to consent for themselves. There was no evidence indicating the need for proxy consent by an authorized representative for any participant according to the UC Davis Alzheimer's Disease Center's procedures for Assessing Capacity to Consent [87]. Each participant received a $20 Amazon gift card for every hour of their time. Given the complexity and the variability in technical skill of people with dementia, we expected that there could be a wide range in the duration of the research sessions (as described by [48]). However, we did not find this to be the case, and each participant session ranged from 2 to 3 hours. As task and time performance were not recorded in this study, there was no potential for additional monetary "reward" to take longer. Furthermore, we did not observe any participant behavior that suggested they were trying to increase the study's duration.

4.2 Procedures

All study procedures were approved by the University **Institutional Review Board** (**IRB**) for the protection of our participants at our university. At all times, we were compliant with IRB requirements. The research team conducted each remote usability study using the web-conferencing software tool Zoom$^{\text{TM}}$. When working with people with dementia it is crucial to minimize the amount of new learning required for participation. For this reason, while Zoom$^{\text{TM}}$ is not the perfect platform, we chose to use it because it is the most widely used by people with dementia, as reflected by its use by the major dementia organizations for their peer-to-peer support groups for people with dementia [24]. It is also considered an accessible video conferencing tool [2]. Each study involved two researchers. One led the study while another observed, took notes, and provided additional support. After each usability test, the research pair discussed the study, major findings, and deviations with the rest of the research team. Following consent, we used the Zoom$^{\text{TM}}$ screen-record feature to capture the participant's audio, video, and on-screen actions. Morphic was then downloaded and installed on the participant's computer by a researcher.

The study consisted of two blocks of the same four tasks. The task blocks compared the Windows10 **operating system** (**OS**) against the application being evaluated by the usability test. Participants completed the first task block using the OS and the second task block using Morphic. Before the second block of tasks, a researcher introduced the participant to the software application.

Prior to the first task, participants were given an overview of the usability test format (i.e., a series of tasks on the task list to complete followed by the System of Usability survey and a brief semi-structured interview) and study instructions. The first task involved participants being asked to increase the magnification of the computer system overall. For the second task, participants were supposed to find the addresses of two places, get directions between them, take a screenshot of the directions, and save it to the desktop. During the third task, participants were supposed to change the overall computer's presentation by selecting a dark theme from the ease of use high contrast settings. The final task asked them to find and open a downloaded MS-Word document and then to use Word's accessibility feature, the immersive reading line focus feature, to isolate each line of the text as they verbally read aloud the first paragraph.

Tasks were designed to be representative of how an individual with dementia might modify the accessibility and ease of use of their computer, based on previous work [25]. The tasks varied in complexity and mimicked real-world scenarios when step-by-step directions may not be available. Participants had to use several approaches to complete the tasks. They often had to use more than one program, locate documents, and parse the tasks for context clues to search for content or information. During the second block, they also had to figure out whether Morphic could be used solely or in conjunction with another program to complete the tasks.

After each task block, the participant completed a **System Usability Scale** (**SUS**) survey, which can be a valid measure when comparing two systems [8, 68]. The SUS is a series of 10 statements with each item having a corresponding five-point Likert scale that described the strength of the participant's agreement with the statement (i.e., 5 = Strongly Agree to 1 = Strongly Disagree) [15]. We scored the SUS in accordance with best practices described by Brooke, using the original wording. Prior to administering the SUS, we explained to the participant that "the system" either meant the Windows10 native features in the first block of tasks or Morphic in the second set of task blocks to compare the perceived usability of both systems.

We scored the participant's responses on the recommended scale of 0–100, which despite the 100-point scale is not scored like a percentage [16]. Scores above 70 are considered to be acceptable, between the high 70's to the upper 80's are good, and if they surpass 90, the system's perceived usability can be considered to be excellent [8].

The SUS was followed by a brief semi-structured interview, where participants were asked to reflect on "the system" they used for that task block as well as on the methods used. This had the added value of providing context to their SUS scores. We also asked questions related to how participants were feeling, if they experienced any frustration or anxiety while participating in the usability testing, if they ever felt uncomfortable being observed, and if they wanted more breaks.

4.3 Flexible Methods

Lazar et al., in *Research Methods in Human Computer Interaction*, state "usability testing is flexible and needs to be structured around the activities that are most likely to result in actual changes to the interface" [48]. The variability between people with dementia (i.e., fluctuations hour-to-hour and a longer-term progression of the condition) and their computers (i.e., different versions of OS or software installed) also required flexible methods during the main study. The first four participants had the software application installed (per our planned methods described in the previous paragraph), completing the study on their computer.

When P5's computer did not meet the technology requirements for the study, the idea of updating their software to install Morphic made the participant anxious. P5 completed the study using the Remote Access method, as described below. P6 had similar problems and browser issues. P6 suggested what we call the Modified Think-Aloud method, also described below. Despite trying to minimize any issues with software compatibility by including the system requirements as part of the screening and scheduling process, technical constraints, like the ones described above, suggests that flexible methods may be necessary when working with people with dementia. Situations when flexible methods may be needed include (1) when requiring software installation might exclude those whose computers did not meet the system requirements or those who experienced network issues impacting their ability to download the application or (2) when eliminating the software installation requirement might reduce the potential that their computer could be temporarily changed, reducing their potential anxiety. Flexible study methods were on-the-fly adaptations during the usability testing, without which it may not have been possible to complete the usability test. Usability testing with P5 and P6 inspired both the Remote Access and the Modified Think-Aloud methods.

The Remote Access Method: This method consisted of the researcher sharing their screen. The participant then took remote control of the researcher's computer, which had the software application already installed on it, to complete the usability testing. Researchers instructed participants on how to request remote control using Zoom^{TM}. Once participants had gained remote control, they went through the tasks, independently making decisions on how to complete the task as if they were using their computer.

Modified Think-Aloud Method: For participants who felt uncomfortable taking remote control of the researcher's computer, they could choose the Modified Think-Aloud method. During this method, the researcher shared their screen on a computer with the software application already installed on it. Participants then verbally dictated to the researcher what to select while going through the task.

In the remaining studies, P7–P10 were given the option to conduct the study by either installing the application on their computer or completing the study using either the Remote Access or the Modified Think-Aloud method. The research team also prepared in advance a virtual machine for participants to interact with as a neutral environment (i.e., an empty browser with one tab open, a neutral background with limited desktop icons visible, and files on the virtual machine) should they prefer that method and if technical constraints necessitated the pivot in methods.

4.4 Pilot Study Findings

We conducted pilot studies with five participants with people with mild to moderate dementia in person. As literature on usability best practices involving people with dementia in summative usability testing is very limited, one aim of the pilot studies was to determine the effectiveness of our research methodology and methods. Based on these pilot studies, we discovered three factors that needed to be addressed in the main, remote study:

(1) The importance of the participant's technology setup: PP3's technology setup included a computer with an incompatible OS and multiple monitors that the computer recognized as being turned on despite being powered off. Their setup caused issues installing Morphic and their effective use of Zoom^{TM}. It also highlighted the need for researchers to work with participants to share their screen during software installation. In the main study, the research team asked during scheduling beforehand whether a participant's computer had more than one monitor.

(2) The benefits of having a caregiver on standby: When PP3 and the researchers were unable to troubleshoot the problem together, PP3's spouse helped with logistical issues. An on-site caregiver may be helpful when unexpected events hinder a participant's ability to begin or complete the study. However, not every participant will have someone on-site who can provide help, and caregivers should not be involved in any of the data collection.

(3) Re-wording task descriptions: Several pilot study participants were anxious to know if they had completed the task correctly. Their feedback revealed the need to provide additional keywords and context clues to indicate to the participant when the task was "done." Some tasks were reworded to avoid abstract or metaphorical language, which is particularly difficult for people with dementia to process [30, 50]. Every task included additional line breaks to assist with scanning.

4.5 Methods of Data Analysis

We used thematic analysis to systematically analyze the data. This approach allowed us to focus on the planning, conducting, and evaluating themes we found to be most salient to summative usability testing [13] to obtain a deeper understanding of the specific usability needs within each of these themes. Our current analysis focused on instances affecting the efficacy of remote summative usability testing. It did not include an analysis of usability findings for either the Windows OS or Morphic.

First, we refamiliarized ourselves with the data. Each member of the research team individually re-watched and performed open coding on the screen recordings from the first, post-pilot participant, which also included the feedback, reactions to the methods, and suggestions for improvement gathered from our participants during the semi-structured interview. These codes were then shared with and critiqued by the research team, resulting in a codebook with the three main themes of planning, conducting, and evaluating. Themes included codes, such as experienced issues with their computer environment, task readability, task relatability, difficulties with the web-conferencing tool, exhibited their agency, expressed anxiety, or wanted researcher support.

Once we reached agreement on the codebook themes, two researchers analyzed each of the nine remaining usability test recordings. As additional codes and subthemes appeared, we integrated them into the codebook until we reached data saturation. Saturation occurred after we had reviewed half of the participant data, at which point no new codes emerged. In total, we analyzed 18 hours and 44 minutes' worth of video recorded remote usability tests. The findings below are the result of the lessons learned from this analysis.

5 REMOTE SUMMATIVE USABILITY STUDY FINDINGS

From planning, conducting, and evaluating the full remote summative usability testing, the next sections discuss the various challenges and nuances to consider at each stage, as well as best practices. We also summarize these takeaway lessons.

5.1 Planning and Logistics

When planning remote usability testing with participants with dementia, it is important to consider several factors, including variability of previous technical experience using remote web-conferencing tools, a participant's personal computing environment, as well as task comprehension and presentation.

5.1.1 Remote Web-Conferencing Tool. While all participants with dementia in this study had prior experience with web-conferencing tools, two key issues came up that required troubleshooting on the part of researchers. First, we had difficulty troubleshooting video conferencing issues (e.g., problems finding the share screen icon) because the video conferencing software we use does not show the controls for the videoconferencing software on the screen being shared (i.e., the participant's screen). Second, participants had difficulty with application window management resulting from the web-conferencing software positioning certain windows "over" or "under" other windows, suggesting a difficulty distinguishing or being able to effectively scan content when the windows are layered (i.e., different z-indexes). For example, several participants struggled to find the location of the chat window if more than the Zoom^{TM} window was open (e.g., other windows and software applications) or when Zoom^{TM} automatically minimized itself when they shared their screen. These observations align with previous work that found people with dementia experience difficulty locating information if it's not all presented on one screen [30].

5.1.2 Individual Computing Environments. Participants' familiarity with and preferences for specific devices, software, OS, and other computing environment settings was also an important impact on participation. For some, their primary computing environment was not a computer, but a smartphone or tablet. Both P4 and P10 expressed their preference for their phones. They used their computers for "things [they] know how to navigate through." Learning how to use new devices can be especially difficult for people with dementia [40]. When it was time to take a screenshot, P10 said, "I don't really know how to do that on my computer," demonstrating how to take one using the side buttons on his phone. In a few usability tests, participants encountered different versions of software such as MS-Word than they were used to. After seeing Word 2010, P5 remarked "This is different," and needed extra time to get used to the application before attempting to complete the tasks. Some participants, like P5, were used to different operating systems completely, like MacOS, and explained similar sentiments like “It’s gonna take me a while. I don’t work on Windows.” Given the importance of familiarity to people with dementia, it is important to either configure the task to the user’s desired platform or give individuals time to familiarize themselves with the new system.

5.1.3 *Comprehensibility of Task Descriptions.*

Standard usability testing uses task lists with clear tasks that require no additional explanation from researchers [48]. Although providing a background scenario at the beginning of an entire task list is considered appropriate practice [48], we added background scenarios to each of the specific tasks we asked participants to complete, to avoid creating a need to remember the initial background scenario. We avoided writing long scenarios, so that participants did not miss key contextual clues, which past work notes can be difficult for some people living with dementia to discern [66]. Even though we improved task clarity based on our pilot studies, some participants still had trouble understanding tasks, sometimes interpreting the task scenarios in a way that differed from what researchers intended. All participants misunderstood elements of the task wording during at least one of the tasks. For example, several participants had different interpretations of the definition for “contrast.” P1 physically adjusted the brightness of their monitor, and P7 interpreted the word “contrast” in the literary sense (i.e., compare-and-contrast). Potential ways to minimize this confusion might be to include the task-specific definition for words that cause confusion in pilot studies or to include an image of the end state participants should be working toward. Following the recommendations described in previous research with people with dementia [18, 43, 46, 64], we inserted a line break to facilitate easy scanning as a way of increasing reading comprehension and made the font larger than 14pt to reduce eye strain. While this appeared to facilitate scanning and isolating keywords for some, the impact was not sufficient for others. In these instances, the researcher gave participants the option to have the tasks read aloud. This was beneficial to participants, like P2, who remarked that hearing the tasks helped her to comprehend them better. Be sure to note any different interpretations of words, as this is a considered usability issue with people who have dementia [74], and in the discussion after the task, follow up with the participant to have them explain their interpretation of the nomenclature used. This can point to more usable text content.

5.1.4 *Usability Testing Task List Presentation.*

Usability testing typically includes a list of representative tasks that the user must complete. Usability task lists ensure the participant is directed by a set of clear goals throughout the evaluation [48]. During in-person usability sessions, users will either be given a hard copy of task(s) or the task(s) will be digitally displayed so that the participant may refer to the task whenever necessary. In this remote usability study, the researchers provided digital copies of the task list to participants as a Microsoft Word document.

Every participant had difficulty differentiating between the application that contained the task list and where they should make changes to the various accessibility settings at first. This issue emerged primarily in the first task, where participants would attempt to change the magnification in Microsoft Word, where the task was being displayed, rather than using the computer settings to apply system-wide changes to magnification as described in the task scenario. P1, P4, P7, and P10 attempted to zoom-in to the text in Microsoft Word, P5 increased the browser’s zoom rather than adjust the magnification in the system settings, and P6 increased the magnification in Open Office. During the second task, P9 tried searching for “Miller Senate Office” using the Microsoft Word’s Search toolbar. Similarly, for the third task, both P4 and P7 tried to change the contrast within Word.

As described above, during the first task, participants’ conceptual understanding of what was being evaluated appeared to be limited to and contained within the context of the application window that displayed the task list. Participants appeared to make this initial mental connection because the task list first appeared in Word—they seem to have assumed that the task that they needed to do also occurred within Word. By the fourth task (i.e., the halfway mark), however, every participant understood that the tasks were not limited to the task list application window. This pattern suggests that participants may have shifted their mental models during a usability evaluation of what "the system" being evaluated or the computer is. It may also suggest that the speed in which a mental model recovers and updates for a system or application can differ between participants and from task to task as they grow accustomed to the activity of usability testing.

While researchers did provide an overview of usability testing procedures as instructions prior to the first task, these findings suggest that the first task may require additional clarification of the overall study process. Also, if the usability tasks cannot be displayed within the application or website being evaluated, we suggest making the first task a practice task, so that the participant can get used to switching between applications or windows. It may also help them to better differentiate between what they are evaluating and where the task is being presented.

5.2 Conducting the Remote Usability Tests

Participants were positive overall regarding the remote summative usability testing process. In response to a question about how participants felt doing the study over video/web conferencing rather than in-person, P8 said "I don't know if there would be any difference. I was very comfortable doing what we were doing." The insights, thoughts, and suggestions of participants guided us in shaping the methods, procedures, and suggestions for future improvement. We found that supporting participant's agency, accommodating mental fatigue, and reducing anxiety were key to ensuring participant comfort and facilitating researcher insights. Below, we discuss these strategies.

5.2.1 Supporting Agency. In line with epistemological shifts in technology research with people with dementia [37, 46, 47, 58, 59], one of the main goals of our research was to determine methods for conducting remote summative usability testing that support agency, self-efficacy, and inclusion of people with mild to moderate dementia. We found three opportunities to support agency: (1) always asking permission before making changes in participants systems; (2) letting participants decide whether to keep any changes to their technical environment made by researchers; and (3) reaffirming participants control of whether to continue attempting or moving on to the next task.

Each time we made a change in the participants' system through remote control we first asked participants for their permission, explained what we were doing and why, giving the participant time to take in the information and ask any questions they may have had. We reassured participants that they could stop the researchers at any time if they felt uncomfortable with any changes. This was especially important to support agency while conducting the usability test remotely, where researchers could take full control of participants technical environment. For instance, while interacting with P5 the researchers said, "I won't change without your permission, you can see what I am doing... If you feel uncomfortable, I'll stop. You're in charge here." P5 then asked for alternative options, showing that researchers successfully facilitated her with the opportunity to find a choice that she was comfortable with.

To support agency in usability testing with people with dementia, whether remote or in person, it is the researchers' responsibility to remind participants to undo any change that happened in a participant's system settings while they were conducting tasks. For example, a researcher suggested to P6 to change back the contrast in his environment, but P6 didn't respond and moved on to the next task not indicating he wanted it back to regular contrast. Researchers then brought it back up a minute later when P6 indicated he did in fact want it back to regular contrast, responding: "Yeah. Oh! Yeah. Oh! that's better, right." Researchers should work to minimize any disruption to participant's setting familiarity, as past research in dementia has shown familiarity to be very important [11]. Preserving people with dementia's computer interface settings is equally crucial to usability. In other instances, the agency of participants was supported by keeping the changes made by researchers to their system. For example, P7 liked the brown background with black writing contrast setting and expressed her desire to keep it because it was easier for her to read as opposed to the default contrast settings. When researchers observe a participant struggling with a task, they can remind individuals that it's their choice if they want to continue attempting a task or move ahead to the next task. However, it is essential to not "over-reassure." In one instance, we felt the need to intervene and remind P3 that it was okay to move on when he struggled with a task. But P3 responded and reminded the researcher of a best practice for people with dementia: "Don"t do, be… don't be so quick to jump in and help. I want to be able to try to figure it out by myself. If I need help, I'll ask… I'll tell you." However, other participants wanted regular reassurance.

5.2.2 Accommodating Mental Fatigue. As past research in gerontology has shown, prolonged mental effort for people with dementia can cause mental fatigue, which can negatively affect the ability to concentrate and problem solve [12, 57]. We found mental fatigue affected participants' ability to troubleshoot, improvise, and recover from unexpected situations. Breaks were therefore critical for participants to complete the roughly 2-hour remote study. Even with a break and interview in between blocks of task, individuals described experiencing mental exhaustion when engaging in the tasks. P8 remarked during the second block of tasks, "I thought I had this. I'm stuck here ... I'm running out of gas on you here … I've lost focus." When P7 reached the map task during the second task block, the mental fatigue resulted in their inability to manage the various open windows and tabs. This was the first moment in the study when this participant experienced any issue with either window or tab management. When confronted with task fatigue, participants with dementia often opted to stop the task after being presented with the option. To reduce any unnecessary stress or feelings that may negatively affect feelings of self-efficacy, the option of "pausing" or moving on should be presented throughout both in-person or remote usability studies. Mandated breaks between tasks can also be useful, particularly when the length is determined by the participant—one participant took a couple of hours off between task blocks. Finally, the number of tasks within a single usability session should be carefully limited.

5.2.3 Reducing Participant Anxiety and Frustration. Anxiety is experienced by many people with dementia, "with prevalence estimated from 8% to 71% for symptoms of anxiety and from 5% to 21% for anxiety disorders" [44]. Further, past research indicates that usability testing activities can lead to an increase in negative affect for older individuals [81]. Here, we describe the techniques we used to mitigate the presence of anxiety symptoms. The data on this topic comes from questions we asked during the semi-structured interview following each task block. These techniques were developed to mitigate the presence of anxiety symptoms and are in line with current psychosocial best practices to treat anxiety in dementia [20]. While there are no simple solutions, the following techniques could be helpful at minimizing the presence of anxiety when conducting either in-person or remote usability studies with people with dementia.

First, we let the participant choose when they wanted to conduct the usability study. This allowed them to schedule during the time of day they typically experienced the least fluctuation of their symptoms. We also took great care to allow flexibility based on their needs and fluctuations and adapted accordingly to the situations as they arose, aligning with best practice in usability testing [48].

We observed that different aspects of the usability test could cause symptoms of anxiety at different times with different participants. While conducting the study, most participants indicated that they felt pressure to perform. The pressure they described was self-imposed. Whenever they felt they were not able to perform or complete the task, it would sometimes lead to feeling "embarrassed" [P4] or "ashamed" [P8]. Others felt "irritated" [P2] or "aggravated" [P4], because "I should know this" [P2]. P4 shared, "I was an assistant manager … with over 100 employees and now I can't even do this. It's pretty sad." Participants also expressed they felt pressure to get through the tasks quickly, which may have increased their anxiety. Some participants even apologized when they experienced challenges with tasks.

Whenever a participant either verbally or non-verbally indicated any symptoms of anxiety, we actively worked to reassure, encouraged breaks, and asked if they wanted to move on or stop. Similar strategies were used to lower frustration in usability testing with older adults [28]. Whenever a participant used testing-centric language or engaged in negative self-talk, we reiterated that we were not testing them, rather they were the ones testing the system, as is best practice in usability testing [64]. However, the concept of a test seemed to have a negative connotation for our participants. Participants who verbalized more testing-focused language throughout the study seemed to report more pressure around the completion of the tasks. The focus on testing (and the associated anxiety and frustration) appeared despite our efforts to avoid using testing language, aligning with past research with older adults [28]. This occurrence may stem from the association many people with dementia have with diagnostic practices of physicians to track their cognition. We noted this potential connection during P9's interview when they linked the usability test tasks with the cognitive evaluations their neurologist makes them do.

Researcher prompting and task-scoping: Participants who spent a significant amount of time trying to complete a task without making progress appeared to engage in more negative self-talk, demonstrate less self-efficacy, and lose the ability to recover from unexpected occurrences in tasks. This finding is similar to past work with older adults where lower self-efficacy led to more help-seeking behavior in usability testing [29]. We found when conducting the remote usability tests with the participants who have dementia in this study, these effects can be mitigated in two ways: (1) by encouraging the participant to experiment and (2) by narrowing the scope of the task. When P8 did not know how to start task 3, we prompted him by saying, "A lot of people did not know how to do most of the tasks, but people would kind of play around with the system." When this method did not reduce the effects of stress for participants, they benefited from open-ended questions like, "How do you usually search for information?" In terms of narrowing choices, in his post-study interview, P9 suggested that the researchers could have responded to his inability to locate an icon by saying, "let's look at the buttons on the bottom again. To narrow them down. Ok, so what button do you think we would press to get there? … Let's look at some of the other options we have." While offering a reduced set of choices affects the validity of task time and task completion metrics (discussed in Section 5.3), which are then not representative of everyday use, the benefit of capturing the context of the intervention is twofold. Researchers can note what events necessitated the intervention as well as which prompts were most helpful to remedy the issue. If, however, reducing options does not seem to alleviate the effects of stress and mental fatigue, that is a cue to ask if they would like to continue, take a break, or stop the task and move on to the next task.

Adjusting tesearcher-participant dynamics: The visibility of researchers also affected participant anxiety in some participants. P4 said that, "When I was trying to figure it out knowing that you guys were just watching me. That was a bit unnerving and little stressful because I knew I couldn't get it and I knew you were watching." P2 wanted us to say: "I'm going to work on a couple things, but if you have questions…I'll be here" to reduce the pressure. By doing this, the web-conferencing software could act as a virtual one-way mirror, as is typically used during in-person summative usability studies. Other participants, however, reported less anxiety when researchers were visible. "By seeing you out there, I have more of a comfort level as to who I am talking to" [P9]. Leaving the choice of researcher visibility to participants is one way to support agency and to mitigate participant anxiety in remote usability testing.

Another aspect of researcher-participant dynamics to consider is the traditional lack of conversation during usability testing. Some participants felt the study session should be “more like a conversation.” P3 elaborated further saying, “We [people with dementia] like to be talked to and not talked at.” Participants often appeared anxious when researchers observed them silently. The lack of visual or auditory feedback would prompt remarks from participants like “Are you with me on that?” or concerns of a lost connection. As such, the practice of invisible, silent observation typically employed during usability testing may not be appropriate when conducting remote usability testing with people with dementia. These findings led to the methods adaptation of Modified Think-Aloud, described in the following sub-section.

5.2.4 Participant Reactions to New Methods. Both the Remote Access and Modified Think-Aloud methods appeared to empower participants in all six studies they were utilized. For example, when P7 was provided the option to take remote control of the researcher’s computer she enthusiastically chose the remote access option. She expressed, “That’s the first time I’ve ever remoted into anybody else’s computer… I had a friend from work before who would remote into mine, but I’d never remoted into anybody else’s… I feel important!" Alongside empowering participants, these two novel methods additionally provided context of participant’s technical abilities and flexibility, as well as serving as a source of novel data.

In contrast to previous research with people with dementia [32, 22] and with older adults [28, 63], we found that a think-aloud approach was easy and engaging for participants with dementia. P6’s reaction to Modified Think-Aloud was: “I absolutely love it... My- My temptation would be to make that the way of doing it.” The Modified Think-Aloud approach also led to a better understanding of the usability issues that impact people with dementia. P9 explained, “I felt like I had somebody out there... ’cause if I was going somewhere, she went with me.” Although this method included the participant dictating to the researcher what to do (which affected overall task time), it further developed a collaborative environment facilitating a sense of co-discovery between researcher and participant. Past research conducting usability studies with dyads of older adults found co-discovery to be more effective than think-aloud [28]. In our work, we see the Modified Think-Aloud as combining co-discovery and think-aloud protocol to bring ease to participants with dementia during remote usability testing. We suggest that when possible, all three methods be offered to participants, allowing the participants to choose which method they feel most comfortable.

5.2.5 Remote Access and the Modified Think-Aloud Considerations. Unlike in-person studies, where a standard computer can be prepared in advance with the software being evaluated pre-installed, there is the necessity in remote studies to find alternate means to either support use of the participant’s personal computer, or allow them to use the researcher’s computer remotely. Having the option of choosing which remote method they felt most comfortable with allowed for greater participant comfort, more participant independence, and a wider range of users to participate. Furthermore, by reducing the technical complexity of the study, both methods supported participants to independently make decisions on how to complete the task as if they were using their personal computer.

Both the Remote Access and Modified Think-Aloud methods were beneficial when a participant’s technology environment did not meet system requirements or participants encountered issues with their browsers. These methods were oftentimes workarounds to unresolvable issues that otherwise would have rendered the studies incompletable had they not been implemented. For participants who prefer not to vocalize their choices or prefer to control the cursor, they may prefer the Remote Access method, which allowed them to use the researcher’s computer as if it were their own. Some participants noted that the Modified Think-Aloud method reduced the "testing" atmosphere. For either method, there was also no additional overhead or added time spent in preparation for either protocol from a researcher's point of view.

While participant reactions to the Remote Access and the Modified Think-Aloud methods were overall positive, researchers did observe one drawback to both remote methods when compared to participants using their personal computer. Some participants experienced confusion when they encountered unfamiliar files on the computer or tabs in the browser while using the researcher's computer. If possible, when conducting remote usability testing with people with dementia, researchers should attempt to emulate a familiar computing environment and to avoid unnecessary visual clutter, to reduce participant confusion when using an unfamiliar computer.

5.3 Evaluating Study Results

Promoting full participation of people with dementia in usability testing also requires revisiting established usability metrics. The goal of a summative usability test is typically "to evaluate the effectiveness of specific design choices" using functional interfaces or very high-fidelity prototypes [48]. However, in contrast to previous research on summative usability testing with people with dementia [31], we found that the standard metrics of task and time performance were not the best approaches to evaluate design choices. We observed factors that contributed to a participant's difficulty recovering from different task interpretations, unexpected events, or errors. These factors directly impacted their time, and in some cases, their ability to complete the task. Standard usability metrics, as well as user satisfaction, may need to be modified, interpreted differently, or, in some cases, discarded when evaluating the results of a remote summative usability study with this population. In fact, the first participant in our main study remarked unprompted: "Are you timing me on this? For the comparison? If you are so it's going to be a big problem." Usability testing as a methodology is flexible and responsive to meet the needs of differing users and technologies [48]. It is possible to do summative testing involving people with dementia, even without using the task and time performance metrics, as long as you are using functional software rather than mock-ups or wireframes, and the users are attempting to perform representative tasks. Below, we describe challenges with certain metrics as well as new methods and metrics for evaluations focused on improving the effective use of technology by people with dementia.

5.3.1 Task Performance. We found that the ability to complete tasks is linked not only to the design choices being tested, but other factors, such as emotional activation and visual processing issues. Whenever participants had a different interpretation for a key term in the task or the task contradicted how they would normally approach it (e.g., looking up directions on their computer instead of on their phone), there was a visible increase in participant frustration. This impacted their ability to recover from any differences of interpretation, to pivot when a strategy was not getting the desired result, and ultimately, to complete the task as intended. Another factor affecting task completion was task-driven tunnel vision, or the participants' inability to scan a page to locate the correct area. For example, for task two, P3 had pulled up the map with the directions on the right side of the screen but was scanning the left side looking for it and could not locate the map for several minutes. The participant said that he gave up, even though he had successfully pulled up the map. In this case, measuring completion alone may not properly account for what successful use may look like for this individual. Every participant exhibited this reduced scanning behavior in at least one of their tasks. Challenges in scanning the page may be unique to this population due to the way that dementia affects visual processing. While past research comparing younger and older populations has found differences in time performance but not accuracy, the presence of dementia may affect the utility of task performance as a way to evaluate the results of a usability test [81].

5.3.2 Time Performance. Previous findings concerning the inadequacy of time performance as a metric for usability testing with people with dementia can be found in [11]. As such, we intentionally did not record time performance (task completion time) during the session. Our findings were guided by, support, and further build off of this previous work by providing several additional reasons for this metric's inadequacy. For one, conversation with participants increased the amount of time for tasks without necessarily being linked to the usability of the technology. For the participants who opted to do the usability evaluation using the Modified Think-Aloud method, the back-and-forth nature of the participant telling the researcher what to do, impacted time performance even more. As previously discussed, participants also became fatigued, prompting the need for breaks, affecting completion time. Given the importance of researcher conversation to alleviate anxiety, and breaks to reduce fatigue, we would not advise that researchers refrain from these to preserve the task completion time measure.

The variability of a participant's short-term memory may also impact task completion time. Some participants, like P2, had cognitive carryover because they could recall tasks completed during the first task block. Their ability to remember, learn, and improve on the task the second time around could have shortened the time to complete them in the next task block. P9 used very different searching methods between the first and second block during the directions task. In the second task block, he experienced greater difficulty recovering from unexpected results. Despite being able to complete the same task during the first task block, he was ultimately unable to complete the same task in the second task block. For multiple reasons, time performance does not appear to be an appropriate metric for users with dementia.

5.3.3 SUS. Unlike task performance and time performance metrics, modifications can potentially be made to the SUS to make it more effective in usability testing involving people with dementia. All 15 participants throughout the pilot study and main study completed the SUS. While the researchers verbally went through the paper SUS during the pilot study and the first main study participant, some participants reported having a hard time answering because they could not remember the scale options after the options were read to them. When we went from the pilot study to the fully remote main study, we transitioned from using a paper-based SUS read-aloud to conducting the SUS using Qualtrics online, which was easier for the participants as they could see the scale and did not have to rely solely on memory. Digital settings in Qualtrics did not affect the format of the SUS, and content of the SUS remained the same between the two formats. However, P1 suggested that "[The SUS questions] need to be broken down by each task." He went on to describe that it did not make sense to give an overall option when "some [tasks] were hard, and others were easy," similar to findings from usability studies with older adults [63]. To assist with memory, some participants indicated that they would have preferred doing SUS after each task completed, rather than at the end of the task block. Some participants struggled conceptualizing what the "system" meant in the first question of their SUS survey. Participants #4, #5, #6, and #9 all indicated confusion or wanted clarification as to what "the system" meant. Others responded to the question with a specific task that gave them difficulty in mind. Several participants also verbally provided context to their answers. P5 suggested "on the evaluation questions, a chance for a comment after each question beside [it]. If someone wants to comment… Because sometimes it didn't all fit exactly, ya' know?" While some participants found completion of the SUS challenging due to unfamiliar terminology, these challenges were not caused by the format through which the SUS was administered, whether verbally or online.

Table 2. 10 Lessons Learned

10 Lessons Learned
Multiple overlapping windows in web-conferencing software can be problematic for people with dementia. Try to use web-conferencing software that allows researchers the maximum flexibility to design for simplicity. (5.1.1) ^{a1}
Structure the usability study to allow for tasks to be completed with the user's preferred OS, application, device. Otherwise, allow individuals adequate time to familiarize themselves with the new system. (5.1.2) ^{b1}
Plan the tasks to be as easy to comprehend as possible, by (1) having a short background scenario for each task, (2) using clear, unambiguous language and providing word definitions when needed, (3) using line breaks and font size larger than 14pt, and (4) giving participants the option to have the tasks read aloud to them. (5.1.3) ^{b1}
If the usability tasks cannot be displayed within the application or website being evaluated, make the first task a practice task. The participant can get used to switching between applications or windows. It may also help them to better differentiate between what they are evaluating and where the task is being presented. The first task may require additional clarification of the overall study process. (5.1.4) ^{b1}
Support agency and self-efficacy. Always asking permission before making changes in participant's systems. Let participants decide whether to keep any changes to their technical environment made by researchers. Reaffirm participant control of whether to continue attempting or give up on a task. (5.2.1) ^{b2}
Breaks are critical for participants for any studies lasting over 30 minutes. Otherwise, individuals may experience mental fatigue. Either mandate breaks (and let the participants choose how long the breaks will last), or explicitly offer the option of "pausing" the study or moving on to a new task. (5.2.2) ^{b2}
Reduce participant anxiety and frustration (1) by encouraging the participant to experiment, (2) by the researcher narrowing the scope of the task, (3) by providing hints to the participants, and (4) by giving participants the option for the researchers to be either visible or invisible on their screens, depending on the participant's preference. (5.2.3) ^{b2}
Give participants options so they can decide how they want to participate in the remote study. Let them choose to either (1) install software on their computer, (2) use the Remote Access method, or (3) use the Modified Think-Aloud method. (5.2.4) ^{a2}
If the participant chooses either the Remote Access method or the Modified Think-Aloud where they will interact with a researcher's computer, prepare the computer environment in advance and be sure to reduce any unnecessary visual clutter. (5.2.5) ^{a1}
Task performance and time performance, the standard metrics for usability testing, do not seem to be appropriate for participants with dementia. However, the System Usability Scale, with modifications, may be appropriate for participants with dementia. (5.3.3) ^{b1}

A summary of lessons learned when planning, conducting, and evaluating remote summative usability tests with people who have early-stage dementia. Superscript indicates the lesson learned relates to ^{a}remote usability testing specifically, ^{b}remote and in-person usability testing broadly, ^{1}people with dementia specifically, and ^{2}people with dementia and potentially other people with conditions that affect their cognition or other factors (e.g., age, low self-efficacy) that impact them broadly.

6 LIMITATIONS

The average age of participants was 64. Anyone under the age of 65 who develops any form of dementia is considered to have early onset [1], representing 9% of people diagnosed with dementia, approximately 4,500,000 people world-wide [87]. This relatively younger group of participants may be due to the hesitance of the general population of people with dementia, typically people over the age of 65, to self-identify as a person living with dementia due to associated stigmas [9, 84], making them unwilling to discuss their experiences with researchers [80]. Also, the younger age demographic could also have been due to recruitment criteria of having to use technology on a regular basis and the study moving to remote-only methods (older users may not have felt as comfortable using a technology such as Zoom^{TM}). However, most of this study's participants had either a technical background prior to diagnosis and/or were recruited from an Alzheimer's Association committee that regularly used Zoom^{TM} as a web-conferencing platform.

Another limitation is the lack of ethnic and geographic diversity of the participants. All but two identified as themselves as Caucasian, which is not representative as older black and Hispanic populations see a higher prevalence of dementia per capita [3]. Every participant resided in Western countries (e.g., United States and United Kingdom), which tend toward hyper-cognitivism, or assigning a higher value to cognition over other abilities [7, 39, 72].

There was also a limitation that occurred during the first use of the Remote Access method with P5. This was initially a mistake that was rectified in later sessions. When the research team made the on-the-fly adjustment to the methods due to technical constraints, P5 encountered multiple open tabs in the browser and several files on the background of the desktop that looked "busy" on the researcher's computer that they remotely accessed. The busy-ness appeared to increase the participant's stress levels. In subsequent sessions, the research team prepared a virtual machine in advance for participants to interact with a neutral environment (i.e., an empty browser with one tab open, a neutral background with limited desktop icons visible and files on the virtual machine) should technical constraints necessitate, if they preferred to not alter their personal computer, or if they chose to conduct the usability evaluation using either the Remote Access or Modified Think-Aloud method.

7 FUTURE RESEARCH

Our study revealed several directions for potential future research. First, there is an opportunity to utilize psychometric analysis tools, such as eye-tracking, to investigate the task-driven tunnel vision phenomena exhibited by every participant. Task-driven tunnel vision phenomena could provide new knowledge. How people with dementia scan for information and how scanning is affected when their interpretation of tasks or words differ merit further study. This can result in a more accessible design of information and more effective usability tests.

We also hope to do two comparison studies: (1) We want to compare the remote summative methods with remote formative usability tests, as the outcomes may differ by their implementation during earlier phases of user-centered design. (2) We want to compare in-person and remote summative methods. This future study (when the world health situation allows for in-person studies again) will determine what considerations or modifications may improve in-person summative usability tests with the population.

While the commercial applications we used were serviceable, there is an opportunity to design a usability testing platform for people with dementia that meets their specific needs. The platform could include features to reduce the need to switch between interfaces, multi-tasking, and processes with multiple steps spanning different programs and windows. These design efforts should focus on how to design usability tests that reduce anxiety for participants with dementia.

Though our work revealed considerations when working with established usability methods, there are also new areas for investigation, such as understanding what indicators and metrics are most informative to define the effective use of technology for people with dementia. The limited short-term memory of users with dementia is another usability principle that needs further investigation. It is unclear whether summative usability testing completed over longer periods affects participants' responses and attitudes toward the system. We should also investigate the learning and carryover effects of running usability tests over 2, 3, or 4-day periods. People with dementia may not be able to recall precise details from the evaluation over extended periods. While our study provided considerations for the System of Usability Scale, there is an opportunity to improve its utility as a feedback tool for people with dementia. One such avenue is by using a visual Likert scale similar to those used in studies with people with Down Syndrome [42].

While usability-related data specifics from the Morphic study will be detailed in later work, future researchers interested in involving people with dementia independent of caregivers in their usability studies now may find value in recording the same usability-related metrics that we did. The first usability-related metric was the task's completion status. This status records if the participant successfully completed the task as intended; it also records if there was a failure to complete the entire task (i.e., an impartial completion where the participant completed some, but not the task in entirety); if there was a failure to complete the task due to misinterpretation of the task; or if tasks were left incomplete due to the participant deciding to end the task and move on. For non-success status, it can be helpful to record the contextual factors that contributed to either how they misinterpreted the task or what adversely impacted their ability to complete the task. The qualitative context surrounding a task's failure status can provide helpful details for improving either study methods or the system being evaluated by describing potential points where a user may make a misstep.

The next usability-related data point captured the task's interpretation status recorded (i.e., if the participant interpreted the task as it was intended). If a task was misinterpreted, we described how the participant interpreted it. We also recorded the entire navigational path that the participant took throughout the task to compare it with the "Optimal Path" as another data point. For example, we recorded content like "the participant clicked on [UI element], searched in the search bar for '[query],' and clicked on [UI element]." We also included related information-seeking details such as what words they used when searching for content or help, what links they clicked on and where it was located, and what information-seeking methods and platforms did they appear to use [e.g., Google, Yahoo answers, application-specific community forums].

In this work, we report that efficiency-centric metrics of task and time performance were not the best metrics to evaluate design choices. Responding to previous research that raised concerns about the inadequacy of time performance as a metric for usability testing with people with dementia [7], we supplement their findings with two additional reasons. These are (1) the conversational nature of people with mild to moderate dementia during usability tests and (2) mental fatigue necessitating breaks. Given the importance of researcher conversation to alleviate anxiety and breaks to minimize mental fatigue, we suggest refraining from the use of time performance measures. To expand upon our finding in this work of the ineffectiveness of time-related data, our future work will also include details to both quantify and describe the factors that can distort the effectiveness of the task and time performance as a metric with this population. To do so, we will record the various factors and their impact on time by detailing how much time a participant spent referring to help features or documentation; how much time they spent recovering from unexpected results, behaviors, or events; or if they appeared to be intuitively exploring, how long they experienced technical difficulties, had miscellaneous conversations, and how much time was spent on task verification. We will also include various qualitative measures from verbal interactions or observations, such as when they make task carryover comments or comments that indicate mental fatigue.

In another usability-related data point, we detailed observable help-seeking behavior. This included the total number of times the participant referred to the task as a reminder. The total number of interventions was categorized into four types. Of the four kinds of interventions, type 1 interventions record the breaking of a repeating sequence by asking participants to read the task again or consider another option. Type 2 interventions are when researchers provided a general hint, such as letting participants know the information that they need is on a screen they have been to already. Type 3 interventions provided a specific hint, such as telling participants that the function they need is under a menu item or on the next screen. Type 4 interventions occur when a researcher generally tells a participant what to do or, when all else fails, telling participants to click on a certain button or navigate to an element. An intervention does not include clarifying task instructions or repeating the task, prompting a participant to think aloud, or helping a participant recover from a software bug, computer crash, or some other technical issue. For every intervention type that occurred, we also detailed the context which necessitated the intervention and the intervention method that was used.

The resulting usability-related data included quantitative usability metrics and supplemental qualitative details to provide greater nuance to the data. Recording both quantitative and qualitative usability-related data can be used in a variety of ways, such as ensuring that future summative usability study research is using effective metrics developed from data involving people with dementia independent of caregivers in summative usability studies. While other populations may be able to draw upon large amounts of data that support the validity and effectiveness of typical usability metrics, the lack of independent involvement of people with dementia in summative usability testing points to the necessity of providing additional context to determine. This data will drive the determination of which metrics are definitively effective or ineffective with people who have mild to moderate dementia when taking part in summative usability studies.

Finally, even with our recommendations, no task will ever be perfectly worded or usability evaluation devoid of participant stress. However, researchers still have an inadequate understanding of what makes an accessible interface for people with dementia and what aspects create confusion. Future research into this area may yield meaningful contributions to the research literature and lead to the creation of a more inclusive design.

8 CONCLUSION

As discussed in Section 3.2, the existing literature on the efficacy of summative usability testing with people with dementia is limited and provides conflicting results [7, 22]. Our study is the first to conduct remote summative usability testing with people with dementia. Some of our findings, such as providing the option of "pausing" or moving on to accommodate mental fatigue and researcher prompting participants to experiment are relevant to usability testing with people with dementia, whether in person or remote. Asking permission before making changes through remote control of someone's device and leaving the choice of researcher visibility up to participants are relevant to remote usability testing with any population. Other findings such as allowing for flexibility in the method based on participant needs and fluctuation abilities and letting users decide to keep or discard changes made to their personal computer systems during the study relate to remote usability testing with people with dementia. Our work builds on previous work [7], highlighting the importance of preserving people with dementia's computer interface settings. For this population, familiarity is essential to usability.

Our study demonstrates that people with dementia can and should be included in remote summative usability testing. However, careful planning and flexibility are requisite. In addition to yielding feedback on technologies that may increase accessibility and adoption, involving people with dementia in research can add invaluable insights to the field of HCI. When they experienced challenges, individuals proposed new approaches for summative usability testing that may benefit other populations, using the Modified Think-Aloud approach to reach a more representative and international sample of participants; many (1) who may not have as fast or stable of an internet connection to support installation downloads, (2) whose distance may otherwise prohibit their ability to participate in in-person only research, or (3) who may lack confidence when interacting with technology. As remote usability testing will continue to a great extent in the future, if not become more common, we believe our findings relating to adaptations and best practices for remote summative usability testing methods may be adapted or extensible to other hard-to-recruit populations.

REFERENCES

[1] Alzheimer's Association. 2020. Younger/Early-Onset Alzheimer's. *Alzheimer's Disease and Dementia.* Retrieved March 16, 2020 from https://alz.org/alzheimers-dementia/what-is-alzheimers/younger-early-onset.

[2] AccessSIGCHI. 2020. Accessible remote attendance. *AccessSIGCHI.* Retrieved July 14, 2020 from https://accesssigchi.com/accessible-remote-attendance/.

[3] Alzheimer's Association. 2019. 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* 15, 3 (2019), 321–387. Retrieved from https://www.alz.org/media/documents/alzheimers-facts-and-figures-2019-r.pdf.

[4] Morten Sieker Andreasen, Henrik Villemann Nielsen, Simon Ormholt Schrøder, and Jan Stage. 2007. What happened to remote usability testing? An empirical study of three methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*, 1405–1414. https://doi.org/10.1145/1240624.1240838

[5] Cathrine Arntzen, Torhild Holthe, and Rita Jentoft. 2016. Tracing the successful incorporation of assistive technology into everyday life for younger people with dementia and family carers. *Dementia (London, England)* 15, 4 (2016), 646–662. https://doi.org/10.1177/1471301214532263

[6] Arlene J. Astell, Maggie Ellis, Lauren Bernardi, Norman Alm, Richard Dye, Gary Gowans, and Jim Campbell. 2007. Developing technology to support the relationship between people with dementia and caregivers. In *Proceedings of the 22nd Conference of Alzheimer's Disease International*, 30–33.

[7] James C. Ballenger. 2000. *Concepts of Alzheimer Disease: Biological, Clinical, and Cultural Perspectives.* Johns Hopkins University Press, Baltimore and London.

[8] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction* 24, 6 (2008), 574–594. https://doi.org/10.1080/10447310802205776

[9] Susan Behuniak. 2011. The living dead? The construction of people with Alzheimer's disease as zombies. *Ageing and Society* 31 (2011), 70–92. https://doi.org/10.1017/S0144686X10000693

[10] Jeffrey P. Bigham and Anna C. Cavender. 2009. Evaluating existing audio CAPTCHAs and an interface optimized for non-visual use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*, 1829–1838. https://doi.org/10.1145/1518701.1518983

[11] Jennifer Boger, Babak Taati, and Alex Mihailidis. 2016. Interdisciplinary development of manual and automated product usability assessments for older adults with dementia: Lessons learned. *Disability & Rehabilitation: Assistive Technology* 11, 7 (2016), 581–587. https://doi.org/10.3109/17483107.2015.1063714

[12] Ann Bossen. 2010. The importance of getting back to nature for people with dementia. *Journal of Gerontological Nursing* 36, 2 (2010), 17–22.

[13] Virginia Braun and Victoria Clark. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[14] Margot Brereton, Laurianne Sitbon, Muhammad Haziq Lim Abdullah, Mark Vanderberg, and Stewart Koplick. 2015. Design after design to bridge between people living with cognitive or sensory impairments, their friends and proxies. *CoDesign* 11, 1 (2015), 4–20. https://doi.org/10.1080/15710882.2015.1009471

[15] John Brooke. 1996. SUS: A "quick and dirty" usability scale. In *Usability Evaluation in Industry*, Patrick W. Jordan, B. Thomas, Ian Lyall McClelland, and Bernard Weerdmeeste (Eds.). Taylor & Francis, 189–204.

[16] John Brooke. 2013. SUS: A retrospective. *Journal of Usability Studies* 8, 2 (2013), 29–40.

[17] Anders Bruun, Peter Gull, Lene Hofmeister, and Jan Stage. 2009. Let your users do the testing: A comparison of three remote asynchronous usability testing methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*, 1619–1628. https://doi.org/10.1145/1518701.1518948

[18] William P. Burns, Chris D. Nugent, Paul J. McCullagh, Huiru. Zheng, Dewar D. Finlay, Richard J. Davies, Mark P. Donnelly, and Norman D. Black. 2008. Personalisation and configuration of assistive technologies. In *2008 30th Annual*

International Conference of the IEEE Engineering in Medicine and Biology Society, 3304–3307. https://doi.org/10.1109/IEMBS.2008.4649911

[19] Kapil Chalil Madathil and Joel S. Greenstein. 2011. Synchronous remote usability testing: A new approach facilitated by virtual worlds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11),* 2225–2234. https://doi.org/10.1145/1978942.1979267

[20] Georgina Charlesworth, Susan Sadek, Astrid Schepers, and Aimee Spector. 2015. Cognitive behavior therapy for anxiety in people with dementia: A clinician guideline for a person-centered approach. *Behavior Modification* 39, 3 (2015) 390–412. https://doi.org/10.1177/0145445514561317

[21] B. Cook and P. Twidle. 2016. Increasing awareness of Alzheimer's disease through a mobile game. In *2016 International Conference on Interactive Technologies and Games (ITAG'16)*, 55–60. https://doi.org/10.1109/iTAG.2016.16

[22] R. J. Davies, C. D. Nugent, M. P. Donnelly, M. Hettinga, F. J. Meiland, F. Moelaert, M. D. Mulvenna, J. E. Bengtsson, D. Craig, and R.-M. Dröes. 2009. A user driven approach to develop a cognitive prosthetic to address the unmet needs of people with mild dementia. *Pervasive and Mobile Computing* 5, 3 (2009), 253–267. https://doi.org/10.1016/j.pmcj.2008.07.002

[23] Marianne Dee and Vicki L. Hanson. 2016. A pool of representative users for accessibility research: Seeing through the eyes of the users. *ACM Transactions on Accessible Computing* 8, 1 (2016), 4:1–4:31. https://doi.org/10.1145/2845088

[24] Dementia Alliance International (DAI). 2017. Dementia alliance international peer-to-peer support groups. *Dementia Alliance International.* Retrieved June 13, 2020 from https://www.dementiaallianceinternational.org/services/online-support-groups/.

[25] Emma Dixon and Amanda Lazar. 2020. The Role of sensory changes in everyday technology use by people with mild to moderate dementia. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'20)*, 1–12. https://doi.org/10.1145/3373625.3417000

[26] Susan Dray and David Siegel. 2004. Remote possibilities? International usability testing at a distance. *Interactions* 11, 2 (2004) 10–17. https://doi.org/10.1145/971258.971264

[27] Yuan Feng, Suihuai Yu, Dirk van de Mortel, Emilia Barakova, Matthias Rauterberg, and Jun Hu. 2018. Closer to nature: Multi-sensory engagement in interactive nature experience for seniors with dementia. In *Proceedings of the 6th International Symposium of Chinese CHI–(Chinese CHI'18)*, 49–56. https://doi.org/10.1145/3202667.3202674

[28] Rachel L. Franz, Ron Baecker, and Khai N. Truong. 2018. "I knew that, I was just testing you": Understanding older adults' impression management tactics during usability studies. *ACM Transactions on Accessible Computing* 11, 3 (2018), 1–23. https://doi.org/10.1145/3226115

[29] Rachel L. Franz, Leah Findlater, Barbara Barbosa Neves, and Jacob O. Wobbrock. 2019. Gender and help seeking by older adults when learning new technologies. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 136–142. https://doi.org/10.1145/3308561.3353807

[30] Ed Freeman, Linda Clare, Nada Savitch, Lindsay Royan, Rachael Litherland, and Margot Lindsay. 2005. Improving website accessibility for people with early-stage dementia: A preliminary investigation. *Aging & Mental Health* 9, 5 (2005), 442–448. https://doi.org/10.1080/13607860500142838

[31] Aideen Gibson, Claire McCauley, Maurice Mulvenna, Assumpta Ryan, Liz Laird, Kevin Curran, Brendan Bunting, Finola Ferry, and Raymond Bond. 2016. Assessing usability testing for people living with dementia. In *Proceedings of the 4th Workshop on ICTs for improving Patients Rehabilitation Research Techniques (REHAB'16)*, 25–31. https://doi.org/10.1145/3051488.3051492

[32] Grant Gibson, Claire Dickinson, Katie Brittain, and Louise Robinson. 2015. The everyday use of assistive technology by people with dementia and their family carers: A qualitative study. *BMC Geriatrics* 15, 1 (2015), 1–10. https://doi.org/10.1186/s12877-015-0091-3

[33] Grant Gibson, Claire Dickinson, Katie Brittain, and Louise Robinson. 2019. Personalisation, customisation and bricolage: How people with dementia and their families make assistive technology work for them. *Ageing & Society* 39, 11 (2019), 2502–2519. https://doi.org/10.1017/S0144686X18000661

[34] Elizabeth Hanson, Lennart Magnusson, Helene Arvidsson, Anette Claesson, John Keady, and Mike Nolan. 2007. Working together with persons with early stage dementia and their family members to design a user-friendly technology-based support service. *Dementia* 6, 3 (2007), 411–434. https://doi.org/10.1177/1471301207081572

[35] H. Rex Hartson, José C. Castillo, John Kelso, and Wayne C. Neale. 1996. Remote evaluation: The network as an extension of the usability laboratory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'96)*, 228–235. https://doi.org/10.1145/238386.238511

[36] Niels Hendriks, Karin Slegers, and Pieter Duysburgh. 2015. Codesign with people living with cognitive or sensory impairments: A case for method stories and uniqueness. *CoDesign* 11, 1 (2015), 70–82. https://doi.org/10.1080/15710882.2015.1020316

[37] Shawn Lawton Henry. 2007. Just ask: Integrating accessibility throughout design. Retrieved on 6 April, 2021 from www.uiAccess.com/JustAsk/.

[38] Maarten Houben, Rens Brankaert, Saskia Bakker, Gail Kenning, Inge Bongers, and Berry Eggen. 2019. Foregrounding everyday sounds in dementia. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS'19)*, 71–83. https://doi.org/10.1145/3322276.3322287

[39] Julian C. Hughes. 2011. *Thinking Through Dementia*. Oxford University Press.

[40] Jaakko Hyry, Goshiro Yamamoto, and Petri Pulli. 2011. Requirements guideline of assistive technology for people suffering from dementia. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL'11)*, 1–5. https://doi.org/10.1145/2093698.2093737

[41] Peter Klein and Martina Uhlig. 2016. Interactive memories: Technology-aided reminiscence therapy for people with dementia. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA'16)*, 1–2. https://doi.org/10.1145/2910674.2935838

[42] Libby Kumin, Jonathan Lazar, Jinjuan Heidi Feng, Brian Wentz, and Nnanna Ekedebe. 2012. A usability evaluation of workplace-related tasks on a multi-touch tablet computer by adults with down syndrome. *Journal of Usability Studies* 7, 4 (2012), 118–142.

[43] Noriaki Kuwahara, Shinji Abe, Kiyoshi Yasuda, and Kazuhiro Kuwabara. 2006. Networked reminiscence therapy for individuals with dementia by using photo and video sharing. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility (Assets'06)*, 125. https://doi.org/10.1145/1168987.1169010

[44] Yong Tae Kwak, YoungSoon Yang, and Min-Seong Koo. 2017. Anxiety in dementia. *Dementia and Neurocognitive Disorders* 16, 2 (2017), 33–39. https://doi.org/10.12779/dnd.2017.16.2.33

[45] Amanda Lazar and Emma E. Dixon. 2019. Safe enough to share: Setting the dementia agenda online. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW: 1–23. https://doi.org/10.1145/3359187

[46] Amanda Lazar, Caroline Edasis, and Anne Marie Piper. 2017. A critical lens on dementia and design in HCI. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI'17)*, 2175–2188. https://doi.org/10.1145/3025453.3025522

[47] Amanda Lazar, Austin L. Toombs, Kellie Morrissey, Gail Kenning, Jennifer Boger, and Rens Brankaert. 2018. HCIxDementia Workshop: Engaging people living with dementia. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA'18)*, 1–7. https://doi.org/10.1145/3170427.3170613

[48] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction*. Elsevier Science & Technology.

[49] Clayton Lewis. 2007. Simplicity in cognitive assistive technology: A framework and agenda for research. *Universal Access in the Information Society* 5, 4 (2007), 351–361. https://doi.org/10.1007/s10209-006-0063-7

[50] Stephen Lindsay, Katie Brittain, Daniel Jackson, Cassim Ladha, Karim Ladha, and Patrick Olivier. 2012. Empathy, participatory design and people with dementia. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems (CHI'12)*, 521. https://doi.org/10.1145/2207676.2207749

[51] Galina Madjaroff and Helena Mentis. 2017. Narratives of older adults with mild cognitive impairment and their caregivers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'17)*, 140–149. https://doi.org/10.1145/3132525.3132554

[52] Juan F. Maestre, Haley MacLeod, Ciabhan L. Connelly, Julia C. Dunbar, Jordan Beck, Katie A. Siek, and Patrick C. Shih. 2018. Defining through expansion: Conducting asynchronous remote communities (ARC) research with stigmatized groups. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18)*, 1–13. https://doi.org/10.1145/3173574.3174131

[53] C. Malinowsky, O. Almkvist, A. Kottorp, and L. Nygård. 2010. Ability to manage everyday technology: A comparison of persons with dementia or mild cognitive impairment and older adults without cognitive impairment. *Disability and Rehabilitation: Assistive Technology* 5, 6 (2010), 462–469. https://doi.org/10.3109/17483107.2010.496098

[54] Shima Mehrabian, Jocelyne Extra, Ya-Huei Wu, Maribel Pino, Latchezar Traykov, and Anne-Sophie Rigaud. 2014. The perceptions of cognitively impaired patients and their caregivers of a home telecare system. *Medical Devices (Auckland, N.Z.)* 8 (2014), 21–29. https://doi.org/10.2147/MDER.S70520

[55] Franka Meiland, Anthea Innes, Gail Mountain, Louise Robinson, Henriëtte van der Roest, J. Antonio García-Casal, Dianne Gove, Jochen René Thyrian, Shirley Evans, Rose-Marie Dröes, Fiona Kelly, Alexander Kurz, Dympna Casey, Dorota Szcześniak, Tom Dening, Michael P. Craven, Marijke Span, Heike Felzmann, Magda Tsolaki, and Manuel Franco-Martin. 2017. Technologies to support community-dwelling persons with dementia: A position paper on issues regarding development, usability, effectiveness and cost-effectiveness, deployment, and ethics. *JMIR Rehabilitation and Assistive Technologies* 16, 4 (2017), e1. https://doi.org/10.2196/rehab.6376

[56] Helena M. Mentis, Galina Madjaroff, and Aaron K. Massey. 2019. Upside and downside risk in online security for older adults with mild cognitive impairment. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*, 1–13. https://doi.org/10.1145/3290605.3300573

[57] Keith Diaz Moore. 2007. Restorative dementia gardens. *Journal of Housing for the Elderly* 21, 1–2 (2007), 73–88. https://doi.org/10.1300/J081v21n01_05

[58] Kellie Morrissey, Amanda Lazar, Jennifer Boger, and Austin Toombs. 2017. HCIxDementia Workshop: The role of technology and design in dementia. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA'17)*, 484–491. https://doi.org/10.1145/3027063.3027083

[59] Kellie Morrissey, John McCarthy, and Nadia Pantidi. 2017. The value of experience-centred design approaches in dementia research contexts. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI'17)*, 1326–1338. https://doi.org/10.1145/3025453.3025527

[60] Zachary Munn, Micah D. J. Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology* 18, 1 (2018), 143. https://doi.org/10.1186/s12874-018-0611-x

[61] National Institute on Aging. 2017. What are the signs of alzheimer's disease? Retrieved on 27 July, 2020 from https://www.nia.nih.gov/health/what-are-signs-alzheimers-disease.

[62] Timothy Neate, Aikaterini Bourazeri, Abi Roper, Simone Stumpf, and Stephanie Wilson. 2019. Co-created personas: Engaging and empowering users with diverse needs within the design process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*, 1–12. https://doi.org/10.1145/3290605.3300880

[63] Barbara Barbosa Neves, Rachel L. Franz, Cosmin Munteanu, Ronald Baecker, and Mags Ngo. 2015. "My hand doesn't listen to me!": Adoption and evaluation of a communication technology for the 'oldest old.' In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1593–1602. https://doi.org/10.1145/2702123.2702430

[64] Jakob Nielsen. 1994. *Usability Engineering*. Morgan Kaufmann, San Francisco, CA.

[65] Chris D. Nugent, Richard J. Davies, Mark P. Donnelly, Josef Hallberg, Mossaab Hariz, David Craig, Franka Meiland, Ferial Moelaert, Johan E. Bengtsson, Stefan Savenstedt, Maurice Mulvenna, and Rose-Marie Droes. 2008. The development of personalised cognitive prosthetics. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 787–790. https://doi.org/10.1109/IEMBS.2008.4649270

[66] Louise Nygård and S. Starkhammar. 2007. The use of everyday technology by people with dementia living alone: Mapping out the difficulties. *Aging & mental health* 11 (2007), 144–55. https://doi.org/10.1080/13607860600844168

[67] Roger Orpwood, Sidsel Bjørneby, Inger Hagen, Outi Mäki, Richard Faulkner, and Päivi Topo. 2004. User involvement in dementia product development. *Dementia* 3, 3 (2004), 263–279. https://doi.org/10.1177/1471301204045160

[68] S. Camille Peres, Tri Pham, and Ronald Phillips. 2013. Validation of the system usability scale (SUS): SUS in the wild. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57, 1 (2013), 192–196. https://doi.org/10.1177/1541931213571043

[69] R. C. Petersen, J. C. Stevens, M. Ganguli, E. G. Tangalos, J. L. Cummings, and S. T. DeKosky. 2001. Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 56, 9 (2001), 1133–1142. https://doi.org/10.1212/WNL.56.9.1133

[70] Helen Petrie, Fraser Hamilton, Neil King, and Pete Pavan. 2006. Remote usability evaluations with disabled people. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'06)*, 1133–1141. https://doi.org/10.1145/1124772.1124942

[71] Mai T. Pham, Andrijana Rajić, Judy D. Greig, Jan M. Sargeant, Andrew Papadopoulos, and Scott A. McEwen. 2014. A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Research Synthesis Methods* 5, 4 (2014), 371–385. https://doi.org/10.1002/jrsm.1123

[72] Stephen G. Post. 2000. *The Moral Challenge of Alzheimer Disease: Ethical Issues from Diagnosis to Dying*. The John Hopkins University Press, Baltimore and London.

[73] Alexandra Queirós, Margarida Cerqueira, Ana Isabel Martins, Anabela G. Silva, Joaquim Alvarelhão, António Teixeira, and Nelson Pacheco Rocha. 2014. ICF inspired personas to improve development for usability and accessibility in ambient assisted living. *Procedia Computer Science* 27 (2014), 409–418. https://doi.org/10.1016/j.procs.2014.02.045

[74] Philippa Riley, Norman Alm, and Alan Newell. 2009. An interactive tool to promote musical creativity in people with dementia. *Computers in Human Behavior* 25, 3 (2009), 599–608. https://doi.org/10.1016/j.chb.2008.08.014

[75] Abi Roper, Ian Davey, Stephanie Wilson, Timothy Neate, Jane Marshall, and Brian Grellmann. 2018. Usability testing—An aphasia perspective. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'18)*, 102–106. https://doi.org/10.1145/3234695.3241481

[76] Dagoberto Cruz Sandoval and Jesús Favela. 2017. Co-designing ambient-assisted interventions using digital interlocutors for people with dementia. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers (UbiComp'17)*, 813–821. https://doi.org/10.1145/3123024.3125615

[77] Marcia J. Scherer, Stefano Federici, Lorenza Tiberio, Martina Pigliautile, Fabrizio Corradi, and Fabio Meloni. 2012. ICF core set for matching older adults with dementia and technology. *Ageing International* 37, 4 (2012), 414–440. https://doi.org/10.1007/s12126-010-9093-9

[78] Jerry Schnepp and Brent Shiver. 2011. Improving deaf accessibility in remote usability testing. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'11)*, 255–256. https://doi.org/10.1145/2049536.2049594

[79] Panote Siriaraya and Chee Siang Ang. 2014. Recreating living experiences from past memories through virtual worlds for people with dementia. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI'14)*, 3977–3986. https://doi.org/10.1145/2556288.2557035

[80] Karin Slegers, Andrea Wilkinson, and Niels Hendriks. 2013. Active collaboration in healthcare design: Participatory design to develop a dementia care app. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems (CHI EA'13)*, 475–480. https://doi.org/10.1145/2468356.2468440

[81] Andreas Sonderegger, Sven Schmutz, and Juergen Sauer. 2016. The influence of age in usability testing. *Applied Ergonomics* 52 (2016), 291–300. https://doi.org/10.1016/j.apergo.2015.06.012

[82] Marijke Span, Marike Hettinga, Myrra Vernooij-Dassen, Jan Eefsting, and Carolien Smits. 2013. Involving people with dementia in the development of supportive IT applications: A systematic review. *Ageing Research Reviews* 12, 2 (2013), 535–551. https://doi.org/10.1016/j.arr.2013.01.002

[83] Neeraja Subrahmaniyan, D. Jeffery Higginbotham, and Ann M. Bisantz. 2018. Using personas to support augmentative alternative communication device design: A validation and evaluation study. *International Journal of Human–Computer Interaction* 34, 1 (2018), 84–97. https://doi.org/10.1080/10447318.2017.1330802

[84] Kate Swaffer. 2014. Dementia: Stigma, language, and dementia-friendly. *Dementia* 13, 6 (2014), 709–716. https://doi.org/10.1177/1471301214548143

[85] Päivi Topo and Britt Östlund. 2009. Dementia, Design and Technology: Time to Get Involved. IOS Press.

[86] Thomas Tullis, Stan Fleischman, Michelle Mcnulty, Carrie Cianchette, and Marguerite Bergel. 2002. An empirical comparison of lab and remote usability testing of Web sites. Retrieved on 3 March, 2021 from https://www.researchgate.net/publication/228540469_An_empirical_comparison_of_lab_and_remote_usability_testing_of_Web_sites.

[87] UC Davis Alzheimer's Disease Center. 2002. Policy and procedures for assessing capacity to consent for research. Retrieved on 29 July, 2020 from https://health.ucdavis.edu/clinicaltrials/StudyTools/Documents/ResearchCapacityPolicy7=02.pdf.

[88] Gregg Vanderheiden, Jutta Treviranus, José Angel Martinez Usero, Evangelos Bekiaris, Maria Gemou, and Amrish Chourasia. 2012. Auto-personalization: Theory, practice and cross-platform implementation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56 (2012), 926–930. https://doi.org/10.1177/1071181312561193

[89] ISO/TR 16982:2002. ISO. Retrieved 2021 from https://www.iso.org/standard/31176.html.

Received July 2020; revised April 2021; accepted April 2021