Assessing Usability Testing for People Living with Dementia

Assessing Usability Testing for People Living with Dementia

Aideen Gibson
School of Nursing, Ulster University
Magee Campus
Derry/Londonderry BT48 7JL, UK
+44 28 71675221
a.gibson@ulster.ac.uk

Claire McCauley
School of Nursing, Ulster University
Magee Campus
Derry/Londonderry BT48 7JL, UK
+44 28 71675293
c.mccauley2@ulster.ac.uk

Maurice Mulvenna
School of Computing & Mathematics,
Ulster University
Newtownabbey BT37 0QB, UK
+44 28 90368602
md.mulvenna@ulster.ac.uk

Assumpta Ryan
School of Nursing, Ulster University
Magee Campus
Derry/Londonderry BT48 7JL, UK
+44 28 71675350
aa.ryan@ulster.ac.uk

Liz Laird
School of Nursing, Ulster University
Magee Campus
Derry/Londonderry BT48 7JL, UK
+44 28 71675006
ea.laird@ulster.ac.uk

Kevin Curran
School of Computing & Intelligent
Systems, Ulster University
Derry/Londonderry BT48 7JL, UK
+44 28 71675565
kj.curran@ulster.ac.uk

Brendan Bunting
School of Psychology, Ulster
University
Coleraine Campus
Coleraine BT52 1SA, UK
+44 28 7167 5295
bp.bunting@ulster.ac.uk

Finola Ferry
The Bamford Centre, Ulster University
Belfast Campus
Belfast BT15 1ED, UK
+44 28 7167 5088
f.ferry@ulster.ac.uk

Raymond Bond
School of Computing & Mathematics,
Ulster University
Newtownabbey BT37 0QB, UK
+44 28 90368156
rb.bond@ulster.ac.uk

ABSTRACT

This paper reports on a study that examines the value of several common usability testing protocols, methods and metrics when used to evaluate the usability of a new personalised reminiscence 'app'. The app, called 'InspireD', is a bespoke app designed to support personalised reminiscence for people living with dementia. The study focused on determining the value of commonly used methods for evaluating usability of apps designed for use by people living with dementia and their caregivers. The study indicated that observation and recording of task completion rates and times produced the most reliable results. The think-aloud methodology was difficult for the people living with dementia and did not produce any reliable data. Thinking-aloud whilst doing a task may have been a distraction since it requires a higher cognitive load. The systematic usability scale score which is derived from a post-test instrument is not reliable, as it had no association with the task completion times.

REHAB 2016, 13-14 October 2016, Lisbon, Portugal.

Keywords

Usability protocols, usability, user experience, UX, human computer interaction, reminiscence, apps.

1. INTRODUCTION

Dementia is a progressive condition, for which there is currently no cure. The World Health Organisation estimates that there are approximately 47.5 million people living with dementia worldwide and there are 7.7 million new cases diagnosed every year [1]. It is estimated that in the UK there are 850,000 people with a diagnosis of dementia [2]. Dementia affects memory, thinking, language, judgment, and it ultimately affects the way a person communicates. For people living with dementia, their ability to present rational ideas and to reason lucidly is diminished [3]. However, it has been demonstrated that people living with dementia can participate in research and provide useful feedback on Information Technology (IT) solutions [4]. The development of new treatments for dementia has become a UK government priority [5]. Alongside research for effective treatments of dementia, there is an urgent need for research, innovation and developments in therapeutic interventions that provide immediate and much needed support to transform the care and lives of people as they live with dementia.

Reminiscence is an activity that can enrich the lives of people living with dementia. Reminiscence is the sharing of memories relating to personal life experiences. The act of reminiscing can serve many functions that create bonds between people and in doing so, supports them to reflect on important life events and to attribute meaning to their lives [6]. The use of traditional prompts aimed at stimulating feelings and memories can be supported by the application of reminiscence systems to help this activity [7]. Reminiscence systems have been defined as 'the use of technology to support reminiscence work' [8]. Technology that facilitates reminiscence increases opportunities for people living with dementia to participate in conversations and to enhance their social interactions [9].

Many existing software systems, apps and online social networking websites provide the capability to gather, browse and share multimedia resources. However, there is very little research into the usability of these systems for the purpose of reminiscing amongst people with deteriorating cognitive function. In 2012, Thiry [10] discovered that many older people do not use social networking sites or online communities because there is 'too much going on'. Their research indicated a need for software systems which are 'simpler and minimalistic, offering only the most basic support for content creation and management.'

The need to involve all stakeholders in the system design and to undertake usability testing of the user interface is imperative and this is widely accepted as good practice [11, 12]. As a result, human-computer-interaction researchers have proposed standard instruments, protocols and metrics for measuring 'usability' as a construct [13, 14]. However, where the target user group has diminished cognitive abilities and perhaps also physical impairments, issues can arise that pose problems when using these standard methods for usability testing [15].

As we move towards an inclusive society and the use of computer applications or 'apps' and ubiquitous devices become an integral part of everyday existence, there is an implicit need to design digital systems that can be used by all, regardless of their physical or cognitive abilities or impairments. It is therefore important that the design and development of digital systems and apps, should formally involve the intended target user group, and that their contribution should be evidenced in usability protocols. This paper describes the development and usability assessment of the InspireD app, which is a reminiscence system for people living with dementia. It proposes that in order to make user involvement a success there is a need to select traditional usability protocols carefully and tailor the evaluation/ testing sessions to suit the target user group.

2. 'InspireD' – A Reminiscence app

The two primary aims of the app are to enable people living with dementia and their family caregivers to select and store personalised memorabilia (photographs, videos, sounds, music) and to provide easy access to these visual and audio-visual cues to support bespoke reminiscence. An Agile software development approach [15] was adopted to allow a functional prototype to be created early in the development lifecycle. The design is minimalist, using verbal descriptors as well as images and icons to reinforce and indicate functionality to the user.

3. Evaluating Usability

Usability is measured in terms of how easily a system can achieve its goals and how efficiently a user can interact with the system through its user interface. Nielson defines usability as 'a quality attribute that assesses how easy user interfaces are to use' [17]. Standard protocols to measure these attributes can be classified as: observation; concurrent thinking-aloud; single ease questions; recording by video and/or audio; and the systematic usability scale which is a post-test survey. These methods in turn provide metrics that can be used by researchers to determine the usability of the user interface.

3.1 Observational approaches

Neilson believes that observing people using a system is the best way to understand what works and what does not work during the user experience (UX) [17]. He advocates a protocol of realistic, representative actionable tasks and the observations of users as they attempt each task to the best of their abilities. The scenarios involve typical tasks that reflect the system's intended use, and these mimic the real world as much as possible.

The concurrent 'Think-aloud' protocol (TAP) is a common observational technique for eliciting insight into the user's cognition and thought processes. It was first utilised for evaluating user interface design by Lewis [18]. This protocol requires the user to perform a number of tasks while 'thinking aloud'. The researcher records the user actions (written or sometimes using tape recordings or video recordings) for each of the tasks, as well as noting any problems and user perplexities.

Video analysis recording (REC) is commonly used to record and measure UX and usability. The availability of small mobile testing units to record user interactions with an app or website can provide invaluable insights into the usability of a system. This moderated 'lab' usability testing scenario is still one of the best ways to capture the rich experience of interacting with a mobile device [19]. It allows researchers to capture the interactions between the user and the device as well as any verbalisation from 'thinking-aloud'.

3.2 Questionnaire-based approaches

The Single Ease Question (SEQ) is a 7-point rating scale to assess how difficult users find a task [20]. Using the 7-point rating scale, the user estimates the level of difficulty of the task before and after attempting it. This measure has greater validity since the metric is recorded immediately after each task as opposed to the end of the session.

The systematic usability scale (SUS) is a post-test survey, first used in 1986, that has become an industry standard questionnaire for measuring the usability of a system [21]. It consists of 10 questions which facilitate answers in a Likert scale format. Each question has 5 response options (or ratings between 1 and 5 where 5 = strongly agree). The systematic usability scale instrument is a well-balanced survey since it consists of 5 questions with negative connotations and 5 with positive connotations. All Likert ratings are then converted to a systematic usability scale score (or SUS score) and the mean SUS score is used to represent the usability of the system. A mean SUS score greater than 68 is considered above average since this is the accepted mean SUS score from a distribution of SUS scores previously collected from usability tests.

3.3 Task completion-based approaches

Task completion rate (TCR) is the percentage of users who completed the task [22]. Task completion is probably the most important metric that determines the usability of the system. For example, if a user cannot accomplish a representative task using a system, then that system is poorly designed. Thus, a 100% task completion rate is the objective for any system since its intended purpose should be intuitive to its user base. The inverse of this metric is the task failure rate.

Task completion time (TCT) is the amount of time in seconds required by a user to complete a given task [22]. An associated metric is the time-until failure, which is the amount of time a user is willing to dedicate before giving up on completing the task.

4. Study Design

The aim of this study was to explore and assess the value of usability protocols for a reminiscence app in the context of use by people living with dementia and their caregivers. Together with 7 dyads, with each dyad comprising a person living with dementia and their primary caregiver, researchers investigated the appropriateness, validity and reliability of several common usability tests and matrices. The tests and matrices for investigation in this study comprise: Concurrent think-aloud protocol (TAP) [18]; Video recording and audio recording devices (REC) [18,19]; Task completion rates (TCR) [21]; Task completion times (TCT) [21]; Single Ease Questions (SEQ) [19]; and Systematic Usability Scale (SUS) [22]. The value of the tests and metrics was examined in a series of five workshops during a 6-week period. The study was approved by the Research Ethics Committee of Ulster University.

5. Experiments

A user group was established to test the usability of the app for people living with dementia and their caregivers. The group comprised of 7 adults living with dementia (5 men and 2 women) and 7 caregivers (all women). One of the dyads was involved from the outset of the study to inform its design, and early stages in the development of the app. This dyad constitutes the lead user dyad (LU). Additional participants were recruited from the Alzheimer's Society Home Support Network. Adults interested in participating were provided with detailed information about the study, and given opportunities to clarify their role in the planned workshops and to discuss any concerns about the study with members of the research team, before the formal consenting process. The age range of participants (n=14) was 42 - 77.

The first workshop (DLU) consisted of a pilot test which was conducted in the home of the lead user dyad to identify any potential issues or barriers that may arise for people living with dementia when testing the app. The subsequent 4 usability workshops were undertaken with the other six dyads over a period of 2 weeks. Two of the workshops (GW1 and GW2) were managed as a group intervention in the university, and the remaining two workshops (DW1 and DW2) were managed as individual dyad interventions, undertake in their homes. In addition to the workshops, the dyads tested the utility of the app over a period of 4 weeks at home.

5.1 Preparation

The first (GW1) and last (GW2) group workshops were conducted in the form of an introductory group meeting and a final focus group, respectively. Dyads participated in the two usability workshops (DW1, DW2) in their own home, where each person living with dementia and each caregiver was instructed to perform a series of tasks using the app while being observed by the researchers who watched and took notes. These tasks, for example, to 'Open Music folder and find the song by The Beach Boys' were scheduled to be completed on two separate occasions with each person being observed for around thirty minutes. It was planned that the first set of tasks would be recorded using an audio recording device and the second set of tasks would be recorded using a video recording camera.

Twelve typical tasks were identified for the users to complete in DW1 and DW2. These tasks were carefully written so they would be realistic, actionable and avoid unnecessary prompting from the caregiver or the researchers. The tasks in DW1 related to using the app to do simple reminiscing – interacting with photographs, watching movie clips and listening to audio clips. The tasks in DW2 concerned selecting, uploading and recording materials to the app to use for reminiscing.

Table 1 - Engagement matrix with usability measures

	DLU	GW1	DW1	DW2	GW2
TAP	✓		✓	✓	
REC			✓	✓	
TCR			✓	✓	
TCT				✓	
SEQ		✓			
SUS					✓

Table 1 illustrates the range of usability protocols and metrics and their employment with the lead user couple (DLU), group workshops (GWn) and dyad workshops (DWn). The following sub-sections discuss outline the protocols used with these different groupings of participants.

5.2 Lead User Dyad Workshop (DLU)

The lead user dyad comprised a man, living with dementia (pseudonym 'Mike' for the purpose of this paper) and his female caregiver. Both individuals tested the app during a 1-hour session to establish whether there would be any potential issues that would prevent people living with dementia from participating in the usability workshops. Mike was asked to complete a set of defined tasks and his interactions with the app were observed and recorded so that the researchers could establish a protocol for the user development workshops. Input and opinions from his caregiver were also recorded and these helped in the planning and preparation of the workshops for the user development group. Mike was aged 42 at the time of testing and had a high level of computing skills. He also had his own collections of digital photographs, videos and music.

Usability testing consisted of 'think-aloud' task analysis where they described what they were doing and their thinking process behind each interaction. Mike was asked to comment on the image quality, display and size of text on the user interface and the sound. Feedback was also sought concerning the size of buttons and the help button and features were commented on. Observations were made regarding how easy/difficult it was for Mike to interact with the touchscreen device. No obvious barriers were identified that would prevent people living with dementia from testing the app.

5.3 First Group Workshop (GW1)

Sauro and Lewis [21] argue that product and domain experience have much more impact on usability metrics than demographics. Accordingly, in GW1, the participants were introduced to the function of the tablets (iPads) in order to minimise a digital literacy bias. They were not introduced to the InspireD app at his stage of the study, in case this might influence their ability to complete the usability tests in DW1 and DW2.

After meeting and greeting participants, researchers explained the purpose of GW1 and its context within the greater research study. Consent forms were signed and questions from the participants were answered. The pre-test questionnaire was given to all the participants to determine their previous experience and use of IT systems.

The participants were provided with verbal instructions on how to turn on an iPad, launch an app (Safari), close the app and turn the device off. Participants were then encouraged to look at generic photographs on Flickr and were shown how to 'swipe' from one page to the next.

5.4 First Dyad Workshop (DW1)

This workshop was undertaken with each dyad in their home. The purpose of DW1 was to evaluate the usability of the InspireD app as an aid to reminiscing. Each participant, irrespective of whether she/he was a person living with dementia or a caregiver, were given the same tasks, and their task completion rates and task completion times were recorded.

The six tasks for completion pertained to using the app to support simple reminiscing. The tasks involved looking at photographs, watching movie clips and listening to audio clips. The participants were asked to estimate the level of difficulty of each task using SEQ before and after they attempted it. The researcher recorded their responses on a grid.

5.5 Second Dyad Workshop (DW2)

The second home-based usability workshop aimed to evaluate the usability of the app in terms of ease in adding users to the system and uploading reminiscence materials, e.g., photographs, videos and audio clips. On this occasion, the task completion rates and task completion times were recorded. Researchers took notes at all interviews and a mobile observation device (MOD-1000, a USB macro camera) was used to record the image of the participant's tablet while it rested on the table. This small device is mounted on a lightweight aluminium plate with a grip-tight surface and its size would suggest that it can be used unobtrusively to observe the use of the app. Similarly to DW1, a task completion grid was completed by one researcher while the other researcher took notes of observations and issues raised. Participants were asked to estimate the level of difficulty of each task using SEQ before and after they attempted it, and the researcher recorded their responses on a grid.

5.6 Second Group Workshop (GW2)

After using the tablet device at home for a period of 1-2 weeks, the user group reconvened as a focus group within the university setting to evaluate the experience of using the app. Both positive and negative feedback was recorded at this meeting to give as much insight as possible into overall user satisfaction. The focus group was recorded using an audio recording device. All of the participants were asked to complete the SUS survey to measure the groups' perceptions of the usability of the app.

6. Results

No obvious barriers to interaction with the touch screen device were identified during DLU testing with the lead user dyad, which could prevent people living with dementia from testing the app. Mike indicated that the image quality, display and size of text on the user interface and the sound quality were satisfactory. Whilst he was able to use most of the buttons easily, the Help and Exit buttons posed some problems. Mike experienced some difficulty relating to 'thinking aloud'. He strayed off topic and could not describe the actions he was carrying out or what he was

thinking as he attempted to complete the tasks. His caregiver had to bring him back to the actual task and steer the conversation towards the app.

6.1 General findings

The data generated in the group and dyad workshops were analysed in a similar participatory approach to that used of Brankaert et al. [23]. In DW1, it was evident that all of the caregivers could interact comfortably with the app when using it to browse reminiscing materials. The task completion rate for the caregivers was 100%. Task 2 (Scrolling through a group of images) presented challenges for all but one of the participants living with dementia. Only two of the participants living with dementia were able to complete task 6 (going back to the previous screen and exiting the app). All of the other tasks were completed by at least three of the participants living with dementia. It is noteworthy that one participant living with dementia was unable to complete any of the tasks. In DW2, the tasks were completed in pairs comprising the person living with dementia and his/her caregiver. It was found that 96% of the tasks in DW2 were successfully completed in this mutually supportive approach.

The researchers had planned to use the think-aloud data to illuminate the experience of using the app. Think aloud was piloted in DLU. It became apparent in the workshops that participants living with dementia had difficulty in verbalising and narrating what they were doing, even when prompted and reminded to do so during completion of the tasks.

In DW2, the MOD-1000 mobile observation device was used to record the image of the participant's tablet. Its intended use was to record each dyad uploading materials to the app. It became apparent after only 15 minutes that the device presented a distraction that was interfering with the ability to complete the specified tasks. Mobile observation is a tried and tested method for measuring usability. In short, the participants perceived that the MOD-1000 was part of the reminiscence system. Consequently, a decision was taken in our study, not to use the camera device in the subsequent workshops.

The task completion times for DW2 varied slightly depending on the age and experience of the participants. Researchers had estimated that it would take approximately 30 minutes to complete all 6 tasks. The participant that identified himself as most experienced in the use of IT systems completed all 6 tasks in 25 minutes. In contrast, the slowest completion rate was 34 minutes.

The difficulty ratings in the 6 tasks completed as a mutually supportive dyad in DW2 were recorded and analysed (see Table 2).

Table 2 - Results of SEQ for DW2

Task	Expected difficulty rating (edr)	Actual difficulty rating (adr)	Delta (edr - adr)	p-values
1	4.17 (2.93)	3.00	-0.03	0.59
2	3.50 (2.88)	3.67	-1.02	0.92
3	3.33 (2.94)	2.67	-2.13	0.59
4	3.67 (2.58)	2.83	-1.25	0.86
5	2.33 (1.03)	2.67	0.33	0.58
6	3.50 (2.43)	3.83	0.05	0.47

A negative Delta value indicates the task was easier than expected, while a positive value means that the task was harder than anticipated by the user. In this study, although most of the tasks were actually easier than the user dyads had anticipated they would be, the results are insignificant given the small number of user dyads in the study (see p-values in Table 2).

The systematic usability scale for post-test survey has become an industry standard questionnaire for measuring perceptions of usability. The mean rating given to the InspireD app by caregivers was 67.5% (SD=11.55) and the 4 people living with dementia who completed the SUS questionnaire awarded the app 78.75%. These results indicate that the app is usable, as a mean SUS score greater than 68 is considered above average [21]. However, the task completion rates (TCR) observed indicate that the app was more usable for caregivers than for people living with dementia. This challenges the widely accepted reliability and validity of the SUS methodology of measuring usability. The most plausible reason for these discrepancies is that the participants living with dementia had a different perception of difficulty than that of the caregivers. Their replies indicated that they enjoyed using the app, that they would recommend it to a friend and that it was a pleasant experience. However, it is possible that they found the questions in SUS difficult to understand or perhaps they could not fully recall the issues they encountered when using the app, after the task. This has implications for future research, given the understanding that short term memory is likely to be adversely affected in many of the common dementias.

7. Discussion

The study indicated that observation and recording of task completion rates and times produced the most reliable results, while the think-aloud methodology was very difficult for people living with dementia, and did not produce any reliable data (Table 3). People living with dementia also found it difficult to assign a value for the SEQ (pre- and post-task ratings). Asking them to assign a number to a perceived difficulty rating was confusing and only the caregivers were able to give a reliable difficulty level to these questions. It was also found that completing post-test questionnaires administered after an event presented a difficulty for the participants living with dementia, perhaps exposing short-term memory loss. As a result, the reliability of the SUS scores could not be

assured. The overall SUS rating given to the InspireD app by caregivers was 67.5% and the 4 people living with dementia who completed the SUS questionnaire awarded the app 78.75%. However, the task completion rates indicated that caregivers found it easier to use the app, than people living with dementia.

In general, post-test surveys such as the SUS instrument can be difficult since they require accurate retrospective reflection of their user experience and the SUS survey itself has an intricate design where the Likert scale of each question alternates between the highest rating being positive and negative feedback. A total of 10 consistent errors/usability issues were identified as a result of the usability evaluations. These were all identified by the researchers observing the participants using the system and were confirmed by the completion rates and the focus group.

The methodology selected to assess usability, the choice of venue to carry out the usability testing and the amount of time given to allow participants to feel comfortable are all likely to have affected the results of the tests.

Table 3 - Summary of findings on the suitability of usability measures

	Summary of findings
TAP	Requires intensive facilitator and/or caregiver interaction and management, supporting prospective memory of person living with dementia
REC	The MOD-1000 camera device was removed as it was found to distracted users when they were completing assigned tasks. In addition, it was perceived by the users to be an additional component of the reminiscence device.
TCR	This was found to be a reliable usability metric for all usability tests independent of user profile.
TCT	This was found to be a reliable usability metric for all usability tests independent of user profile.
SEQ	Not useful for people living with dementia as they find it difficult to estimate how difficult a task should be, perhaps exposing a lack of experience with digital technology.
SUS	The SUS was an invalid instrument in this study. The scores from users living with dementia were not reliable, as they did not concur with task completion rates. This may be because any post-test survey relies on reflection and short term memory. An additional challenge presented to users living with dementia, is alternating negatively and positively worded questions.

In this study, researchers wanted to establish whether using standard tests and matrices is adequate for evaluating the usability of an app where the target users have some form of cognitive impairment. As dementia is likely to adversely affect short-term memory, and the thinking and reasoning functions of the brain, the protocols which involved estimating values, e.g. levels of difficulty or describing processes, e.g. thinking aloud as they completed a task were the most problematic for the users.

8. Conclusion

The InspireD app, created to facilitate the process of reminiscence therapy was tested using standard usability metrics and methods by people living with dementia and their caregiver over a period of approximately 6 weeks. Our research suggests that use of a post-test survey such as SUS may not be reliable when measuring the user experience of people living with dementia since these users suffer from a cognitive condition that affects their short term memory. In addition, we found that the camera based mobile usability testing unit (MOD 1000: Mobile Observation Device) could not be used because it confused the user and caused them to assume that this device was part of the mobile application. Audio recording was also unnecessary since little to no 'think-aloud' data were recorded given that people living with dementia find it difficult to verbalise their human-computer interactions. Our conclusion is that standard protocols used to test the usability of IT systems and apps may not be appropriate for use by people living with dementia. It is not enough to test the usability of a system using protocols where the measurement tools themselves may cause distress or confusion to the system users. Just as it is important to consider the needs of the user when using the system, it is equally important to be aware of the suitability of the criteria we are employing to measure its usability.

In conclusion, our research indicates that with a small sample size the InspireD mobile app is usable for some people living with dementia. Our results showed that people who do not have dementia found the app easy to use and could support people living with dementia to use it to reminisce. The paper also indicates that common usability testing protocols such as the SUS instrument, think-aloud protocols and external mobile macro cameras attached to the mobile testing device may not be suitable for evaluating apps whose target users have been diagnosed with a progressive cognitive disease such as dementia. This suggests that there is a research opportunity to design new protocols or to optimise existing protocols to improve the data collected from usability testing of devices and apps in these contexts.

9. ACKNOWLEDGMENTS

The app was co-created and designed with input from the Alzheimer's Society, Reminiscence Network Northern Ireland (RNNI) and people living with dementia and their caregivers. Special thanks to Alex Turnbull, Software Engineer from Kainos Software Limited for invaluable assistance in technology selection decisions.

REFERENCES

[1] WHO Dementia Factsheet, http://www.who.int/mediacentre/factsheets/fs362/en/, Accessed 11 May 2016.

[2] Alzheimer's Society, Dementia 2014 report statistics, https://www.alzheimers.org.uk/site/scripts/documents_info.php?documentID=341, Accessed 11 May 2016.

[3] NHS, Communicating with people living with dementia, http://www.nhs.uk/conditions/dementia-guide/pages/dementia-and-communication.aspx, Accessed 11 May 2016.

[4] Span M, Hettinga M, Vernooij-Dassen M, Eefsting J, Smits C. (2013) Involving people with dementia in the development of supportive IT applications: a systematic review. Ageing Res Rev., 12(2): 535-51.

[5] Prine Minister's Challenge on Dementia 2020, Department of Health, February 2015.

[6] Butler, R.N. (1963). The life review: an interpretation of reminiscence in the aged. Psychiatry, 26: 65-76.

[7] Gibson, F. (2011). Reminiscence and Life Story Work: a practice guide. London and Philadelphia: Jessica Kingsley Publishers.

[8] Mulvenna, M.D., Astell, A.J., Zheng, H., Wright, T., Reminiscence Systems, In: Mulvenna, M.D., Astell, A.J., Zheng, H., Wright, T., (Eds.), Proceedings of First International Workshop on Reminiscence Systems, Cambridge, UK, September, 2009, pp. 2-4.

[9] Lazar A., Thompson H., Demiris G. (2014). A systematic review of the use of technology for reminiscence therapy. Health Educ. Behav. 41, 51S–61S.

[10] Thiry, E., (2013) Designing a Digital Reminiscing System for Older Adults. SIGACCESS Access. Comput., 105: 24–28. doi:10.1145/2444800.2444805.

[11] Carroll, J. M., Rosson, M. B.: (2007) Participatory design in community informatics. Design Studies 28(3): 243-261.

[12] Muller, M. J. (2002) Participatory Design: the third space in HCI. The Human-Computer Interaction Handbook 4235.6, 1051-1068.

[13] Lewis, J.R., (2006) Usability Testing: Handbook of Human factors and ergonomics. Third Edition. Hoboken (NJ): John Wiley & Sons, Inc.; p. 1275-316.

[14] Sauro, J., 10 Essential Usability Metrics, http://www.measuringu.com/blog/essential-metrics.php, Last accessed 26 June 2016

[15] Astell, A., Alm, N., Gowans, G., Ellis, M., Dye, R., Vaughan, P. (2009) Involving older people with dementia and their carers in designing computer based support systems – some methodological considerations. Universal Access in the Information Society, Vol 8, pp 49-58.

[16] Aydin, M.N., Harmsen, F., Slooten van K., & Stegwee, R.A. (2005) On the Adaptation of An Agile Information Systems Development Method, Journal of Database Management Special issue on Agile Analysis, Design, and Implementation, 16(4): 20-24.

[17] Nielsen, J., (1993) Usability Engineering, 1st ed., Academic Press, London.

[18] Lewis, C. (1982) Using the 'thinking-aloud' method in cognitive interface design. (IBM Research Report RC 9265, 2/17/82). Yorktown Heights, NY: IBM T.J. Watson Research Center.

[19] Sauro, J., How To Conduct A Usability Test On A Mobile Device, http://www.measuringu.com/blog/mobile-usability-test.php, Last accessed 10 May 2016

[20] Sauro J., and Dumas, J.S., (2009) Comparison of three one-question, post-task usability questionnaires. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). ACM, New York, NY, USA, 1599-1608.

[21] Bangor A, Kortum PT, Miller JT. (2008) An empirical evaluation of the System Usability Scale. Int. J. Hum. Comput. Interact. 24(6):574-594.

[22] Sauro J, Lewis JR. Quantifying the user experience: Practical statistics for user research. Elseiver; 2012, Mar 2016.

[23] Brankaert, R., & Ouden, E. Den. (2015). (Re) Design of a mobile interface: Reflections on an in-context evaluation. In proceedings of Participatory Innovation Conference 2015. The Hague, The Netherlands.