

AN INITIAL USABILITY EVALUATION OF SOME WORD-PROCESSING FUNCTIONALITIES WITH THE ELDERLY

Sergio Sayago and Josep Blat

Interactive Technologies Group, Universitat Pompeu Fabra, Barcelona (Spain)

Keywords: Young elderly people, word-processing functionalities, design of questionnaires, usability evaluation.

Abstract: This short paper addresses two key questions about evaluating the usability of word-processing functionalities with the young elderly: (i) which factor (difficulties understanding the terminology, remembering the steps and using the mouse) is the most strongly correlated with the overall usability of some word-processing functionalities?; (ii) when designing a valid usability questionnaire for the elderly, do we need to adapt standard Likert scales? Both questions are answered after running a two-hour MS Word session at an adult school with five elderly people with experience with computers. The preliminary results point out that difficulties remembering the steps and using the mouse have a strong relationship with the overall usability of the word-processing functionalities evaluated. The responses elicited from elderly people are mostly dependent on the visual arrangement (vertical, horizontal) of standard Likert scales. The elderly draw firmly on everyday scales to answer questionnaires. Nevertheless, Likert and everyday scales differ in significant ways. In everyday scales, the elements tend to be arranged vertically. In addition, top elements are usually regarded as the best or most expensive, interesting, etc. These differences turned out to have a strong impact on questionnaire's validity and the adaptation strategy to cope effectively with them. Replacing numbers with adjectives is an effective design solution because adjectives seem to be easier to understand than numbers for the elderly.

1 INTRODUCTION

According to surveys carried out in the UK and Spain ((Goodman et al., 2002), (Larra, 2004)), the elderly use word-processing applications a lot. Some studies have looked into training older people to use word-processing functionalities (Czaja, S. J. and Lee, C., 2003). Nevertheless, our literature review and work reveal that usability evaluations of word-processing functionalities with older people have received scant attention to date, despite being a key aspect to make them more accessible to the growing older population.

The study presented in this short paper is aimed at addressing two key questions about evaluating the usability of word-processing functionalities with older people:

1. *Which factor (difficulties understanding the terminology, using the mouse and remembering the steps) is the most strongly correlated with the*

overall usability of some word-processing functionalities?

Difficulties understanding the terminology, using the mouse and remembering the steps to carry out computer tasks are frequently agreed as accessibility barriers faced by the elderly. Nevertheless, comprehensible terminology, minimum cognitive load and a reduced number of errors are desirable features of usable interfaces for all (Nielsen, 1993).

2. *When designing a valid usability questionnaire for the young elderly, do we need to adapt standard Likert scales?*

Questionnaires are a widely used method to elicit information from people. Current research shows that questionnaires need to be adapted when administered to older people. ((Schwarz and Knäuper, 2000, Schwarz, 2003)) point out that effects of question order decreases with age whereas response order effect increases with age. This is in part due to the fact that answering questionnaires is a cognitive process, which tend to decline when people get older ((Park and Schwarz, 2000)). It has also been found that elderly people tend to select

“don’t know” responses more often than young people. Older people are said to be very cautious and do not draw on contextual information (preceding questions), unlike young and middle-aged people. In spite of these efforts, very little research has been done on Likert scales. Nevertheless, they are frequently used in standard usability questionnaires (e.g.; QUIS) and exploring the requirement of adaptation is therefore worthwhile.

2 OVERVIEW OF THE STUDY

This study was carried out during an ICT course at La Verneda, an adult school in Barcelona (Spain). Elderly people running ICT courses asked us to organize a MS Word session. They often use MS Word in their teaching activities at the school.

Through contextual interviews carried out at the school, we found out that they use MS Word mostly to deal with lists of participants (students) and create user manuals. Hence, we aimed to focus the MS Word session on these aspects. The following word-processing functionalities were evaluated after a two-hour hands-on session where the participants were asked to create a user manual about “how to copy a text from a web page to a Word document”, a frequent task carried out in the courses at La Verneda: (i) creating a table of contents; (ii) saving a word document as an HTML file; (iii) adding headings and footnotes; (iv) inserting a cross-reference and (v) sorting a list of students in alphabetical order.

Five elderly men ranging in age from 64 to 75 and with experience with computers participated in the study. Although this number of users is very small to draw significant conclusions, five users is somehow suggested as a baseline (Nielsen, 1993) in order to identify potential usability problems, which will need to be validated with more users. The participants had been organizing and running ICT courses for the older population at La Verneda for more than 5 years. In addition, they used a wide range of computer applications such as e-mail and MS Word on a daily basis.

An evaluation questionnaire was designed to elicit feedback from the participants on the usability of the word-processing functionalities. The questionnaire was divided into five sections (one section per scenario or task). Qualitative analysis is based on observation-notes taken during the session.

3 ADAPTING STANDARD LIKERT SCALES TO THE YOUNG ELDERLY

In order to identify errors in the formulation of questions and responses (i.e. standard Likert scales) in the questionnaire, a pilot evaluation inspired by cognitive interviewing laboratory techniques was carried out. 7 users participated in this test, 5 elderly adults (2 women; 3 men) and 2 middle-aged people (1 woman; 1 man). None of them took part in the usability evaluation. All the participants were asked to fill in the pilot questionnaire individually and paraphrase (i.e.; to put it into another way) those questions and responses which were difficult for them to understand. Afterwards, they were interviewed individually in order to gather their feedback (e.g.; is there any question did you find difficult to understand?) and to probe their understanding of both questions and responses (e.g.; what do you think this question means?).

The results pointed out that the elderly participants had difficulties understanding the meaning of standard Likert scales. Although we still do not have a full reason for it, they associated 1 with “the best” and 5 with “the worst” independently of the adjectives used in the Likert scales. In the interviews they explained to us that they rely on life experience a lot in order to answer questionnaires. Nevertheless, everyday and Likert scales differ in several ways. In everyday scales, such as the ‘Spanish National football league’ or ‘the top ten richest women’, the elements are usually arranged vertically. In addition, the top elements are usually regarded as the best or the most. These two factors turned out to be key differences as compared to standard Likert scales.

In order to test the impact of these differences on the validity of the responses elicited, a second pilot questionnaire was designed and evaluated with the same users. The questions and responses were the same in both questionnaires. But the responses were arranged vertically in the second. This has a strong effect on the validity of the results: different responses were elicited from the elderly users. For the same question, they selected the number “1” in the first questionnaire (horizontal scale) and the number “5” in the second (vertical scale). Nevertheless, our post-interviews revealed that they aimed to give the same answer in both questionnaires.

With the aim of overcoming validity concerns we had two alternatives: (i) to replace horizontal with vertical scales; (ii) to replace numbers with adjectives. Although both alternatives were brought up by our users during the post-interviews, all of

them expressed a strong preference towards the latter. They insisted on the fact that adjectives are easier to understand than numbers. The following extract is taken from one of the interviews: “*‘difficult’ always means difficult. However, the number 5 can mean different things*”.

In light of their preference towards adjectives, we decided to replace numbers with adjectives in the Likert scales used in the final version of the questionnaire. It was evaluated again with the same users and the responses elicited from both middle-aged and elderly adults were valid (each user was interviewed individually - after they had completed filling in the questionnaire - to confirm their answers).

4 USABILITY EVALUATION OF WORD-PROCESSING FUNCTIONALITIES

Instruments’ consistency and validity are noteworthy concerns of both usability studies and experimental designs. We calculated Cronbach’s coefficient, frequently used to measure questionnaires’ reliability, for each of the five scenarios. The Cronbach’s coefficients ranged from .77 to 1, and averaged .86. As reviewed in (Black, 1999), .86 indicates a high level of internal consistency. With respect to validity, we took the evaluation carried out in (Lin et al., 1997) as a model and tested the hypothesis that “the questionnaire scores show low levels of usability (high scores) when a lot of help is required by the users”. This hypothesis was confirmed ($p < .03$).

It could be thought that the amount of help provided to our users is not a relevant criterion to assess the validity of questionnaires. Elderly people can raise very different types of questions, such as “I’d like to know more about”, which might not necessarily ask for help. Nevertheless, we took advantage of our presence in the session and we only considered in the analysis those questions in which the participants asked us to support them in carrying out the word-processing functionalities.

Next we present the most salient results of the usability test. The Pearson-moment correlation coefficient was calculated for each factor (difficulties using the mouse, remembering the steps and understanding the terminology). As stated in the previous section, numbers were replaced by adjectives in the Likert scales of the final usability questionnaire. Nevertheless, for analysis purposes all the adjectives were mapped to numbers. The values

range from 1 (e.g.; very easy) to 5 (e.g.; very difficult).

4.1 Creating a Table of Contents

On average, creating a table of contents is a difficult task ($M=2.8$; $SD=.83$). Even though our users had problems using the mouse ($M=3.2$; $SD=.83$), difficulties in remembering the steps to create a table of contents shows the highest correlation coefficient ($r_{task1}=.87$) with the overall usability of this task. Our field notes show that all the participants had to repeat the procedure several times (three or four times) until they finally got to remember it. Nevertheless, this strong relationship is not statistically significant ($p > .1$). The terminology was easy to understand ($M=2.2$; $SD=.44$).

4.2 Saving a Word Document as an HTML File

On average, saving a Word document as an HTML file is an easy task ($M=2.2$; $SD=.44$). There is a perfect positive correlation ($r_{task2}=1$) between difficulties remembering the steps to carry out this task and its overall usability. This finding indicates that the more steps to save a Word document as an HTML file, the more difficult for the elderly this task is. It is worth noting difficulties in remembering the steps are rated as the most difficult factor ($M=2.2$; $SD=.89$). According to our observation-based notes, most of the problems experienced by the participants were brought about by the fact that the HTML option in the “Save As” dialog rendered invisible to them. Hence, they had difficulties in remembering where to click on.

4.3 Creating Headings and Footnotes

On average, creating headings and footnotes is an easy task ($M=2.2$; $SD=.83$). Although all the factors analysed were equally rated ($M=2$), there is a significant relationship between difficulties using the mouse and the overall usability of this task ($r_{task3}=.97$; $t(3)=7.74$; $p < .01$). This finding suggests that the more steps/clicks to create headings and footnotes, the more difficult for the elderly this task is. The field notes show that all the participants had problems in scrolling down and up. This was primarily due to precision using the mouse, which the participants lacked despite having experience with computers.

4.4 Creating Cross-references

On average, this task is neither easy nor difficult ($M=2.6$; $SD=.54$). Nevertheless, the users had serious difficulties in all the factors analysed ($M>3$). Difficulties using the mouse and the overall usability of this task are significantly correlated ($r_{task4}=.91$; $t(3)=3.87$; $p<.05$). This finding indicates that the more steps/clicks to create cross-references, the more difficult for the young elderly this task is. However, it is worth noting that difficulties using the mouse were rated as the most difficult factor ($M=3$; $SD=1$).

4.5 Sorting a List in Alphabetical Order

On average, sorting a list in alphabetical order is a very easy task ($M=1.4$; $SD=.89$). All the factors analysed were rated as easy as well. Nevertheless, only difficulties in understanding the terminology and the overall usability of this task are perfectly correlated ($r_{task5}=1$). This finding indicates that the easier to understand the terminology, the easier for the elderly to carry out this task is. Indeed, our field notes show that all the participants suggested using such a clear terminology in the rest of the tasks.

5 DISCUSSION AND CONCLUSIONS

The preliminary results of this paper show that standard Likert scales should be adapted to the special needs of the elderly. It has been found out that the responses elicited from elderly people are mostly dependent on the visual arrangement (vertical, horizontal) of Likert scales. The elderly draw firmly on everyday scales in order to answer usability questionnaires. Nevertheless, both scales differ in significant ways. In everyday scales, elements tend to be arranged vertically. In addition, top elements are usually regarded as the best or the most important, expensive, and so on. These differences have been found to have a huge impact on questionnaire's validity. Replacing numbers with adjectives in standard Likert scales has proven to be an effective solution to cope effectively with the requirements of the elderly.

Both difficulties remembering the steps and using the mouse seem to play a key role in the usability of many word-processing functionalities for the elderly. Although correlation does not necessarily mean causality, and usability tends to be determined by many aspects, this finding suggests

that significant improvements in word-processing functionalities' usability should be achieved by either paying special attention to or focusing only on these aspects.

The results of this study might be difficult to generalize due to the small number of users and word-processing functionalities tested. Nevertheless, we hope the results can contribute to advance the current state-of-the-art in HCI and the elderly. Future studies are needed to both validate and explore in depth the preliminary results presented in this paper. We are currently working on these issues within our ongoing PhD thesis.

ACKNOWLEDGEMENTS

We would like to thank L'Escola d'Adults de la Verneda-St.Martí for their support and collaboration; especially, Ana Burgués, Ana Zafón, MA Serrano and Elisenda Giner. We would also want to thank the reviewers for their useful and inspiring comments.

REFERENCES

- Black, T. R. (1999) *Doing Quantitative Research in the Social Sciences. An Integrated Approach to Research Design, Measurement and Statistics*, SAGE Publications.
- Czaja, S. J. and Lee, C. C. (2003) In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*(Eds, Jacko, J. A. and Sears, A.) Lawrence Erlbaum Associates, pp. 413-428.
- Goodman, J., Syme, A. and Eisma, R. (2002) In *Proceedings Volume 2 of the 16th British HCI Conference*London.
- Larra, R. M. d. (2004) Fundación AUNA, Madrid.
- Nielsen, J. (1993) *Usability Engineering*, Academic Press, Boston.
- Park, D. and Schwarz, N. (Eds.) (2000) *Cognitive Aging: A Primer Aging*, Taylor & Francis.
- Schwarz, N. (2003) *Journal of Consumer Research*, **29**, 588-594.
- Schwarz, N. and Knäuper, B. (2000) In *Cognitive Aging: A Primer*(Eds, Park, D. and Schwarz, N.) Taylor & Francis, pp. 233-253.