

The Impact of Different Icon Sets on the Usability of a Word Processor

T.R. Beelders, P.J. Blignaut, T. McDonald, and E. Dednam

Department of Computer Science and Informatics, University of the Free State, PO Box 339, Bloemfontein, South Africa, 9300

Abstract. This paper discusses the results of usability tests obtained when testing different sets of icons in a word processor environment. An alternative set of icons was developed for a subset of word processor functions and compared to the standard icons. The score obtained for completed tasks as well as the time taken to complete tasks successfully were evaluated. Results indicate that the score is not affected by the icons used in the interface. It was noted that word processor expertise and the icons used have a significant effect on the time taken to complete some tasks. However, each of these factors exhibits an effect in only a single task completed in the prototype. Possible reasons for the significant difference are discussed.

Keywords: Usability, icons, interface.

1 Introduction

The advent of the graphical user interface (GUI) resulted in an increase in the use of icons within computer applications [1]. Users have exhibited distinguishable preferences for interface components such as language, navigation, symbols and colour use [2]. These facts motivate the need for careful consideration of, amongst others, translation and icon development in user interfaces [3] – factors which could have an impact on product usability.

This paper will discuss some of the available literature on icons and usability. An outline of the research methodology that was used will be given, followed by a detailed discussion of the experiment results. Finally, a conclusion, based on the analysis of the results, will be drawn.

1.1 Usability

According to the International Standards Organisation (ISO) standard 9126-1 usability is “the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions”. This definition is further expanded upon in ISO 9241-11 where usability is defined as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [4].

Using the above definitions, four distinct components of usability can be identified, namely effectiveness, efficiency, satisfaction and learnability. These components are defined as follows:

- Effectiveness is how well the user is able to achieve that which must be done by using the system [4] and can be measured in terms of accuracy and completeness [5].
- Efficiency is the amount of resources required to complete the desired task [4].
- Satisfaction is the subjective feeling the user has about using the system [4].
- Learnability measures not only the time taken for a user to become familiarised with the system but also how well the user is able to remember system functionality [5].

Shneiderman [6] lists a set of five measurable objectives that can be measured in order to determine the usability of a product. These measurements of time to learn, speed of performance, rate of errors by users, retention over time and subjective satisfaction allow for specific and controlled evaluation of a software application [6]. The available usability models also provide a number of measurements which can be used by developers to comprehensively test the usability of a product [7].

1.2 Icons

Icons are a common interface component that employ images to represent an object or an action that can be carried out by the user [8]. Continued use of icons has been attributed to the fact that they are easier for users to learn [1], [9] and to use [1]. Their use also increases the productivity of the user since recognition is generally faster for a picture than for text [1], [8].

One disadvantage of icons is that they can easily be misinterpreted by users if the chosen image invokes unintended associations [8] – the picture that “speaks a thousand words may say a thousand different words to different viewers” [10].

Zammit [10] investigated the effectiveness of pictorial and text icons with a group of 11 and 12 year olds and found that neither the pictorial nor the text icons were always immediately recognisable to the users. Users’ accuracy has been shown to be the highest when selecting from a mixed format of text and graphics in a menu structure as opposed to a graphics only or text only menu structure [1]. However, no discernable difference in the time taken to make a selection was detected between the three formats [1].

This research undertook to test the usability of a set of preferred icons chosen by non-computer literate users by means of empirical testing. To complete the set of chosen functions, the remainder of the icons were developed by means of a brainstorming session. By comparing the alternative set of icons to the standard word processor icons, it can be determined whether the alternatives are better suited to South African users and in so doing, establish whether or not there is a need to develop new word processor icons for a South African audience.

2 Research Procedure

2.1 Method

By making use of a simple word processor application the effect of different sets of icons on the usability of a product was tested. Two sets of abstract pictorial icons [10] were used in order to determine if the icons used influence product usability.

2.2 Word Processor Prototype

A small word processor application was developed in order to test the subjects. The word processor possessed minimal capabilities, while still ensuring that it was representative of a fully-fledged word processor or advanced text editor. Functions which were incorporated into the word processor prototype included document handling (e.g. open and close), text formatting (e.g. font size and style) and text manipulation (e.g. copy, cut and paste).

The prototype also allowed for capturing of the users' demographic information, such as age, gender and language.

Users were required to complete a number of small tasks, representative of common word processor tasks. The tasks were displayed sequentially and individually at the bottom of the word processor window (Fig. 1) and could be completed solely by making use of either a toolbar shortcut (icon) or a menu option.

The prototype allowed for real-time evaluation of the tasks as the user completed each one. A number of measurements were also captured for each task. These include the number of menu options and toolbar shortcuts selected by the user, together with the number of keystrokes and mouse clicks and the time required to complete the task.

2.3 Interface

Two sets of icons were used in the different interfaces, namely the standard icons currently found in the Microsoft Office packages and an alternative set of icons obtained from previous studies [11] and via two brainstorming sessions (Fig. 1).

The set of icons obtained during the first brainstorming session were distributed amongst potential word processor users. Given the complete set of icons, respondents were required to indicate which icon they would choose for each of a number of listed word processor functions. All the icons were available to be chosen for each function and icons could be chosen more than once. The alternative icons used in this study for the functions of **Open**, **Close**, **Save**, **Cut**, **Copy** and **Paste**, were chosen as the preferred icon by the group of non-computer literate respondents. There were two icons chosen by the same number of respondents for the **Close** function. Therefore, the icon for **Close** was selected by a process of elimination since one of the icons was also chosen as the preferred **Save** function by a large margin. A number of these icons were confirmed in the same manner in an independent study undertaken by Teklebrhan and Blijnaut [11].

The remainder of the icons were developed during a second brainstorming session and included in the design without confirmation by non-computer literate users. The icons were developed to provide more context for novice and first-time users. For example, the icons used for **Bold**, **Italic** and **Underline** consisted of a bold, italic or underlined capital letter "F" respectively. This was done in an effort to convey to the user the font changes that would occur if the function were invoked. By using the same letter throughout and placing them adjacent to one another on the toolbar, it allows for easier visualisation of the font styling (Fig. 2). It was hoped that by developing in such a manner that novice and first-time users would easily relate to the concepts depicted by the icons.

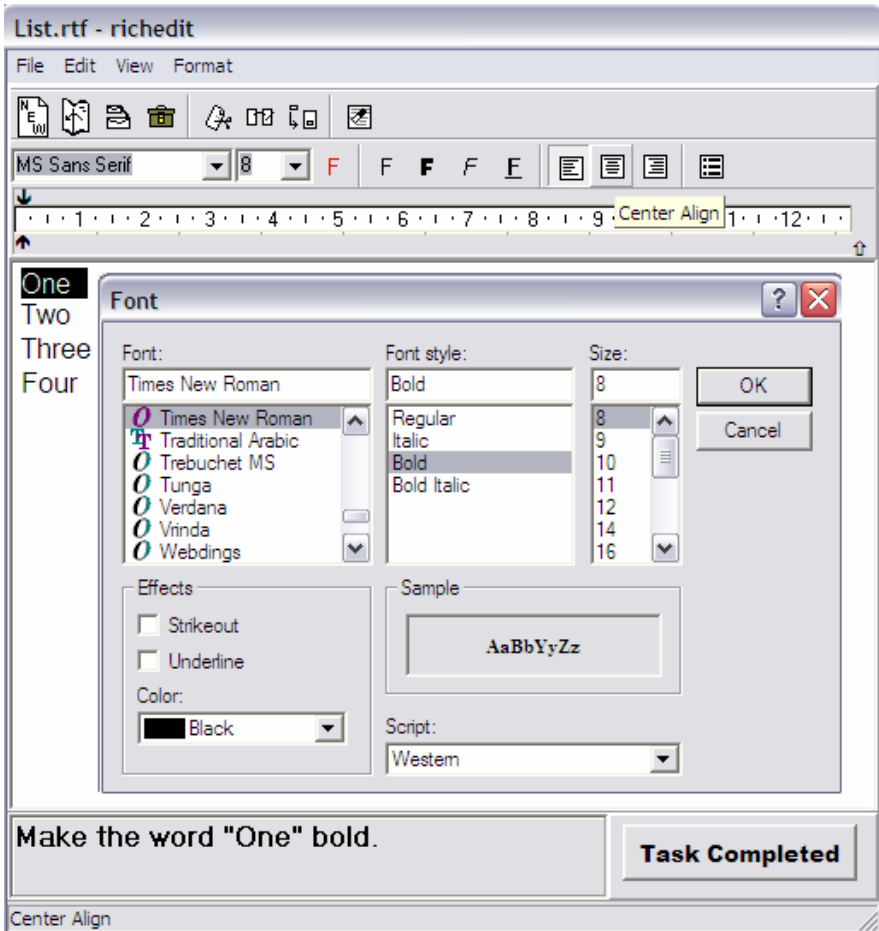


Fig. 1. Word processor prototype with alternative icons

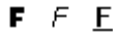


Fig. 2. Font styling icons

To ensure that the effect of the icons could be tested without interference from other interface components, the interface had neither menus nor tooltips. The user had to rely entirely on interpretation of the icon.

In summary, the two interfaces tested were:

- Standard icons with no menu and no tooltips.
- Alternative icons with no menu and no tooltips.

2.4 Subjects

The test subjects consisted of first year university students that were taking a basic computer literacy course. Test subjects spoke a variety of languages, including English, Afrikaans, Sesotho and isiZulu. All subjects were conversant in either English or Afrikaans as these are the tuition languages of the university. The participants provided for different levels of word processor expertise. There were 61 female and 37 male participants who completed the test on either a standard interface or an alternative interface.

2.5 Testing Environment

The test was conducted during the first practical session of the course. This was before the users had received any instruction in word processor packages. Up until that point they had only been taught basic Windows usage.

Each participant was randomly assigned to one of the interface groups. After all the practical sessions had been completed there were 98 tests that had been completed, of which 47 were completed on the interface with standard icons and 51 using the alternative icon interface.

3 Analysis

The usability measures that were analysed were (a) an overall score (discussed below), (b) time taken to complete a task, (c) number of actions required to complete a task, (d) number of errors incurred whilst completing a task, (e) user satisfaction. The number of correct and incorrect answers was also compared for each task. Of these only the effectiveness measurement of score and the efficiency measurement of time [7] will be discussed in this paper.

3.1 Independent Variables

The independent variables used in the analysis were the interface employed during the test and the word processor expertise of the user.

Each user was classified as either a first time or an expert [6] word processor user based on their level of experience with a word processor application, together with the frequency with which they make use of such an application.

3.2 Dependent Variables

The two dependent variables discussed in this paper are the score for each user and the time taken to complete each task. To calculate the score each task was assigned a difficulty index based on the minimum number of actions and inferences required to complete the task successfully. This allowed for a weighted score to be computed for each user. The time taken to complete each task was measured in seconds and then converted to $1/\text{time}$ for further analysis.

3.3 Analysis of Score

The evaluation of the score was done by means of a 2 x 2 between subjects factorial ANOVA. The following hypotheses were formulated for the score:

1. $H_{0,1}$: Word processor expertise has no effect on the score achieved .
2. $H_{0,2}$: The interface used has no effect on the score achieved.

The word processor expertise of the user had no effect on the achieved score since $H_{0,1}$ could not be rejected ($F_{\text{Expertise}}(1, 94) = 0.989, p = 0.322$). $H_{0,2}$ could not be rejected ($F_{\text{Interface}}(1, 94) = 1.192, p = 0.278$), leading to the conclusion that the interface used during the test did not have a significant effect on the achieved score of the user.

3.4 Analysis of Time

The time was evaluated individually for each task by means of a 2 x 2 between subjects factorial ANOVA. Only those tasks that were completed successfully were included in the analysis [12]. The following hypotheses were formulated for the time variable:

1. $H_{0,1}$: Word processor expertise has no effect on the time taken to complete a task successfully .
2. $H_{0,2}$: The interface used has no effect on the time taken to complete a task successfully.

$H_{0,1}$ could be rejected for only one task ($F_{\text{Expertise}}(1, 81) = 4.302, p < 0.05$), where expert users performed significantly better than first time users. The task required users to change the font colour of a word. Two possible explanations could be offered for this difference. Firstly, it was observed during the test that many users experienced difficulty in grasping the concept that the drop-down box containing the font colour can be expanded to reveal a wider selection of colours. Secondly, after changing the font colour, the selection distorts the actual colour of the word. For example, green coloured font appears to be purple when selected. This phenomenon confused users not familiar with the effects that highlighting has on the font appearance. These two observations could possibly have caused some hesitation and confusion on the part of first time users, thus leading to a longer completion time for these users. Evaluation of the number of actions required and the number of errors incurred during completion of the task could provide more information on the cause of the difference.

A second task, which appears slightly later in the test, required users to change the colour of a whole sentence. There was no significant difference exhibited between the users for this second task. This seems to indicate that users retained the knowledge obtained in the previous task and did not experience the same problems again.





Users of the standard icons performed significantly better on the task that required users to close the text document ($F_{\text{Interface}}(1, 45) = 9.797, p < 0.05$), allowing $H_{0,2}$ to be rejected for that task. The alternative icon for the **Close** function was chosen by questionnaire respondents, but the results of this task show that it did not communicate the concept of **Close** as clearly as the standard icon. In fact, the icon chosen by the respondents was actually designed as an alternative for an electronic mail interface. Taking into consideration that choices of non-computer literate users were split evenly between the icon eventually used for **Save** and the one used for **Close**, it may be

pointed out that perhaps the entire concept of closing a document needs to be explained more clearly to novice or first time users.

To place these icons in perspective, they are shown below in Table 1.

The obtained result indicates that although users show a preference for a certain icon, it does not necessarily improve the usability of the product. Icons that are used should be chosen with care and developers should ensure that the icon does indeed convey the intended meaning or concept.

Table 1. Standard and alternative icons

	Standard	Alternative
Save		
Close		

4 Conclusion

The interface had very little effect on the usability of the word processor, a finding which corroborates those of Kacmar and Carey [1] where time is concerned. The only significant difference between the users of the different interface occurred when using an icon that potentially did not convey the meaning of the function clearly to the user. This supports the assertion that careful consideration should be given to the development of icons [3]. The fact that the icon in question was chosen as the preferential icon by questionnaire respondents could indicate a distinct lack of understanding for the concept portrayed by the icon.

Word processor experience only had a significant effect on the task that required use of a complex dialog box, a situation in which it is understandable that a first-time user would show some hesitancy or uncertainty. Subsequent tasks using the same dialog box showed no significant performance difference between users – an indication that users do retain the learned concepts, at least for a short period of time. It would be interesting to test whether users are able to retain this knowledge over a longer period of time than simply between two tasks.

Results would indicate that there is no need for development of an alternate set of icons for South African users. Standard icons appear to be intuitive enough that they correctly convey that which they attempt to represent. Rather, proper explanation of word processing concepts and functions is needed. Given enough time and practice, it appears that most users will be able to master the usage of a word processor application.

References

1. Kacmar, C.J., Carey, J.M.: Assessing the usability of icons in user interfaces. *Behaviour and Information Technology* 10(6), 443–457 (1991)
2. Cyr, D., Trevor-Smith, H.: Localization of web design: An empirical comparison of German, Japanese, and U.S. website characteristics. Retrieved 24 February 2005 from www.loyalty.ca/docs/Localization_of_Web_Design.pdf

3. Johns, S.M.: Colors, buttons, words and culture: Designing software for the global community. Retrieved 24 February 2005 from <http://www.sensi.org/~alec/locale/other/colore~1.htm>
4. ISO9241. ISO 9241-11: Ergonomic requirements for office work with visual display terminals. Beuth, Berlin (1998)
5. Cato, J.: User-centered web design. Addison-Wesley, Great Britain (2001)
6. Shneiderman, B.: Designing the user interface: Strategies for effective human-computer interaction, 3rd edn. Addison-Wesley, United States of America (1998)
7. Abran, A., Khelfi, A., Surya, W., Suryan, W., Seffah, A.: Usability meanings and interpretations in ISO standards. *Software Quality Journal* 11, 325–338 (2003)
8. Benbasat, I., Todd, P.: An experimental investigation of interface design alternatives: icon vs. text and direct manipulation vs. menus. *International Journal of Man-Machine Studies* 38, 369–402 (1993)
9. Schild, W., Power, L.R., Karnaugh, M.: Pictureworld: A concept for future office system. IBM, Thomas J. Watson Research Center, Yorktown Heights, NY, RC 8384 (#36518) (1980)
10. Zammit, K.: Computer icons: a picture says a thousand words or does it? *Journal of Educational Computing Research* 23(2), 217–231 (2000)
11. Teklebrhan, R., Blijnaut, P.: A study on the effect of Western designed metaphors in some culture groups in South Africa. Department of Computer Science and Informatics: University of the Free State, Technical Report 2005/02 (2005)
12. Nielsen, J.: Outliers and luck in user performance. Alertbox (March 6, 2006) http://www.useit.com/alertbox/outlier_performance.html