CONF 921048-6

PNL-SA--20706

DE93 004402

THE EFFECTS OF USING A HYPERTEXT
TOOL FOR SELECTING DESIGN GUIDELINES

Received
DEC 14 1992

J. A. Fox

October 1992

Presented at the
Human Factors Society
Annual Meeting
October 12-16, 1992
Atlanta, Georgia

Prepared for
the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DOLSA-20706

Designing a User-System Interface (USI) is a complex task that has been approached in many ways. One approach has been to use USI design guidelines to help improve the quality and consistency of USIs. To be effective, a general set of guidelines must be tailored to a specific application. One popular set of guidelines is the Smith and Mosier *Guidelines for Designing User Interface Software* (1986). This study investigated the effects of using a computer-based hypertext design aid (DRUID, Dynamic Rules for User Interface Design) for the selection of USI guidelines by both experienced and novice *Guidelines* users.

Results indicate that, in general, DRUID performed as well as the book. Subjects used both the book and hypertext presentation media effectively. However, subjects selected more guidelines that were relevant with the book because subjects read less when using DRUID. Subjectively, the software was preferred because it provided assistance in the selection process and provided additional time-saving design aids not available in the book. Results showed that the presentation of the guidelines could be improved for both media, and many of the factors that led subjects to select fewer relevant guidelines with DRUID could be overcome with improved USI design and implementation of DRUID. This suggests that, to overcome the limitations of presenting large texts on-line, "value added" (e.g. automated checklists) features should be incorporated into on-line tools to help users better perform their tasks.

Hypertext Tool for Selecting Design Guidelines

INTRODUCTION

Good user-system interface (USI) design has become increasingly important as computer systems being designed become more complex and significantly more code is devoted to creating the USI. Numerous models, techniques, and tools have been developed to help the user interface practitioner create, manage, and communicate USI designs.

USI design guidelines are one such tool. Smith (1986) describes guidelines as a series of generally worded recommendations that are used for guidance in the USI design process. Because of their generality, guidelines must be "tailored" to be used effectively for any specific USI (Smith, 1986). Tailoring is the process of creating detailed application-specific USI specifications or rules from the entire set of design guidelines. In this process, a designer selects an initial set of relevant guidelines and then translates them into the design rules (Mosier, 1988).

Some USI designers and researchers have criticized the limited usefulness of guidelines in systems design (Gould and Lewis, 1985; Reaux and Williges, 1988; Thovtrup and Neilsen, 1991; and Woods and Eastman, 1989). Most of the criticisms deal with the user's ability to: (1) understand the meaning and intent of the guidelines and (2) retrieve the applicable subset of relevant guidelines from a larger, often massive, set of guidelines. For example, one of the more comprehensive set of guidelines is the Smith and Mosier *Guidelines for Designing User Interface Software* (1986) (hereafter referred to as the *Guidelines*, which contains 944 guidelines presented over approximately 375 pages.

To help alleviate the problem of accessing and applying guidelines, some practitioners and researchers have called for on-line tools (Normore and Tijerina, 1986; Smith and Mosier, 1984; Perlman, 1989b; Thovtrup and Neilsen, 1991; Tijerina, 1986; Woods and Eastman, 1991). Several hypertext tools have been developed on different platforms. One particular system, DRUID (Dynamic Rules for User Interface Design), allows users to browse the *Guidelines* in a hypertext form. It also allows users to tailor the guidelines and create, store, and print out checklists.

Purpose of this Research

Before guideline tools can be recommended for wide spread use, they must be shown to help the user better utilize the material. This paper presents the results of a study of how USI designers used DRUID to aid in the selection of guidelines from the Smith and Mosier *Guidelines*. Data were collected in an attempt to isolate media-related (hardcopy or softcopy) factors that influenced subjects' performance in selecting guidelines.

METHOD

Subjects

Eight subjects participated in the experiment. All subjects had at least one year of experience in USI design and at least some hands-on experience using the Macintosh computer. Four of the subjects were experienced guideline users who had previously used the *Guidelines* in the design and/or evaluation of a USI. The other four subjects were novices who had, at most, cursory exposure to the *Guidelines*.

Apparatus

An original copy of the 1986 Smith and Mosier *Guidelines* was used as the paper or hardcopy medium. A beta version of DRUID 2.0 was used as the computer-based presentation medium. DRUID provides the users with ability to select guidelines from an abbreviated list of guideline

Hypertext Tool for Selecting Design Guidelines

titles (Figure 1) or a detailed display (Figure 2). For this experiment, DRUID was run on a Macintosh™ IIfx using HyperCard™ software.

Figure 1. Goes here

Figure 2. Goes here

Task Scenario Descriptions

Subjects selected guidelines they thought relevant for each of two system scenarios. The scenarios provided the subjects with the minimum necessary requirements from which a USI could be designed. The scenarios contain information on the users, the task requirements, the environment, and any hardware or software constraints. The two scenarios were designed to be equivalent in terms of level of difficulty in understanding the USI requirements and in the number of applicable guidelines.

Experimental Design

The experiment was conducted as a mixed-factors design with the presentation media (printed *Guidelines* and DRUID) and two similar USI scenarios descriptions as the within-subjects factors. The level of guideline experience (novice or experienced) was the between-subjects factor. Each subject selected guidelines for one scenario per medium.

Guideline Selection Procedure

Phase I. First, a baseline set of necessary or relevant guidelines for the scenarios was developed in order to form the basis for evaluating the subjects' guideline selection in Phase II. The guidelines were independently selected by three guideline experts using DRUID. Guidelines selected by all three experts were considered the "core" set.

Phase II. Prior to selecting guidelines, each subject received training in the use of guidelines and on how the media presents the guideline information. To make the experiment more manageable, the subjects were instructed to select guidelines from only the "Data Display" section of the *Guidelines*. When subjects used DRUID, they were instructed to enter their selections into the software. When using the book, subjects used an unmarked copy of the *Guidelines* and a formatted answer sheet that provided a column to mark selections next to the individual guideline numbers. It should be noted that the answer sheet was created specifically for this experiment to efficiently collect data and that no such sheet is distributed with the *Guidelines*.

Dependent Measures

Four primary measures were taken:

- 1) Percentage of the core guidelines selected. This is known as *recall* in information retrieval literature and is defined as the number of items (guidelines) retrieved and relevant divided by the total number of relevant items in the collection (Salton, 1975).
- 2) Efficiency in selecting guidelines in terms of precision. Precision is defined as the number of relevant items retrieved from a collection divided by the total number of items retrieved from a collection (Salton, 1975).
- 3) Total time to complete the task.
- 4) The number of guidelines selected from the "title" lists.

Hypertext Tool for Selecting Design Guidelines

RESULTS

Phase I

In phase I, the experts agreed on 33 core guidelines for one scenario and 28 guidelines for the other scenario.

Phase II

The measures were evaluated using an analysis of variance (ANOVA). The first primary dependent measure was the percentage of core guidelines selected, or recall. There was a significant difference in performance for presentation medium, $F(1,6) = 25.126$, $p = 0.002$. The recall rate was 91.2% for the book and was 83.4% for DRUID. All eight subjects selected a higher percentage of core guidelines with the book.

The precision measure resulted in no significant differences based on medium. For DRUID the mean precision rate was 34.2% and for the book the mean rate was 39.0%, $F(1,6) = 3.147$, $p = 0.126$. The rates ranged from 19.4% to 50.0%.

There were no significant main effect for the time to complete the task. The mean time when using DRUID was 65.9 and when using the book was 54.12, $F(1,6) = 2.849$, $p = 0.1424$. Pooled across all conditions, the mean time to complete a scenario was 60.02 minutes. The times ranged from a low of 30.43 to a high of 84.87 minutes.

The mean number of guidelines picked from the list was 19.3 with DRUID and was 3.6 with the book, $F(1,6) = 5.483$, $p = 0.058$.

There were no significant interactions. Also, experience level had no significant effect for any measure. The equivalence in difficulty of the two task scenarios was demonstrated by the fact that the scenario main effect was not a significant factor for any measure, as was the intent when scenarios were constructed.

DISCUSSION

Except for the recall measure, the subjects achieved the same high level of performance with DRUID as with the book. The difference in the number of core guidelines selected can be attributed to a variety of factors:

Familiarity with Medium. A problem that must be overcome with any hypertext system is experience: most people have been trained in the effective use of a book since elementary school.

Image Quality and Text Formatting. It has been shown that user performance for on-screen text generally is not as good as displayed on paper (Gould et al., 1987). In this experiment, no adjustments were made for image quality. DRUID was not designed to maximize that aspect of the USI design.

Display Size. DRUID's display area may have also affected the subjects' performance. Hansen and Hass (1988) showed that the size of the screen is a factor in CRT usability. Unfortunately, DRUID was limited to HyperCard's maximum screen size of nine inches diagonal. With the limited screen area, DRUID only displays one guideline at a time. Several subjects commented that they lost context by seeing only one guideline at a time.

Response Time. Another factor was DRUID's response time. It sometimes took DRUID up to three seconds to "turn" between individual guidelines. Most of the subjects negatively commented on how slowly DRUID updated.

Software Errors A bug in the software also influenced several of the subjects' performance. When guidelines were selected from their title lists, the software would sometimes record the wrong guideline selection. Several times the subjects did not notice that they were brought to the wrong guideline, causing them to miss reviewing preceding guidelines.

Selection Strategy. When using DRUID, subjects adopted a selection strategy that reduced the amount of text they read. Whenever possible, the subjects would select the guidelines from just

Hypertext Tool for Selecting Design Guidelines

their titles, instead of reading the guideline itself as they did when the book was used. Also when using DRUID, subjects were less likely to read any annotation provided with the guidelines. Thus, the subjects preferred to select guidelines from the title lists when using DRUID, and from the more detailed guidelines when using the book. However, the time it took to complete the tasks and precision rates were roughly equivalent for both media.

Subjective Ratings. Most of the subjects preferred using DRUID and said that if they had to perform a similar task again they would use DRUID. In general, it seems that it was the software's aids for selecting the guidelines that most contributed to these results. In particular, subjects noted that the software better met their needs and that it helped them keep track of where they were when selecting guidelines.

Effectiveness of Presentation Medium

DRUID provided the experienced as well as the novice users with the design aids necessary to select guidelines effectively. The only measure in which the book significantly outperformed the software was in recall. However, it should also be noted that the experiment was somewhat biased in favor of the book because no formatted answer sheet is distributed with the book.

Fortunately, the software deficiencies that were identified through these results can be alleviated with sufficient design revision. Given such potential improvements, DRUID would be the recommended medium. Not only was DRUID already subjectively preferred by the subjects, its automated checklist features reduce many of the time-consuming tasks associated with guideline selection that go beyond this study.

Until those improvements to DRUID are made; however, potential users of the *Guidelines* should consider the task they need to perform before choosing the medium. For reading, it may be advisable to use the book; for creating checklists, the software. When performing tasks that involve extensive reading and selection, a user may wish to use both media.

Thus, it is possible to produce software tools that can effectively integrate large quantities of text without compromising user performance. Perhaps until the electronic medium can always outperform the hardcopy medium for every task, it is the "value added" aids (such as checklists and report generation) that will determine the success of a hypertext-based system.

Future Research

Further studies need to be performed to evaluate the effectiveness of on-line design aids, such as DRUID, in controlled yet more realistic and complex settings. Real world validity is of utmost importance to the end users of software tools.

The results of this study demonstrate the value of adding job aids to hypertexts to help users perform their tasks. Though DRUID currently provides aids to help users select guidelines, it is easy to envision more sophisticated aids in applying guidelines. In the future, it may be possible to integrate a hypertext-based tool like DRUID with other tools, such as rapid prototyping and decision aids, to help users decide which guidelines are relevant for specific requirements.

DRUID2.0	
DATA ENTRY	Select Functions/Guidelines
1.0 General	Return to 1
	Weight
-1 Data entered only once	Y
-2 Entry via primary display	Y
-3 Feedback during data entry	Y
-4 * Fast response	Y
-5 Single method for entering data	N
-6 Defined display areas for data entry	N
-7 Consistent method for data change	Y
-8 User-paced data entry	I
-9 Explicit ENTER action	---
-10 * ENTER key labeling	---
-11 Explicit CANCEL action	---
-12 Feedback for completion of data entry	---
-13 * Feedback for repetitive data entries	---
-14 * Feedback when changing data	---
-15 Keeping data items short	---
-16 * Partitioning long data items	---
-17 Optional abbreviation	---
-18 * Distinctive abbreviation	---
-19 * Simple abbreviation rule	---
-20 * Minimal exceptions to abbreviation rule	---
-21 * Minimal deviation free abbreviation rule	---
-22 * Fixed abbreviation length	---
-23 * Clarifying unrecognized abbreviations	---

+Opt

Figure 1. Abbreviated Guideline Title List

DRUID2.0	
DATA ENTRY	Select Functions/Guidelines
1.0 General	Return to 1.0
-4 * Fast Response	(4 of 37)
Ensure that the computer will acknowledge data entry actions rapidly, so that users are not slowed or paced by delays in computer response; for normal operation, delays in displayed feedback should not exceed 0.2 seconds.	
Weight: N	
EXAMPLE: A key press should be followed by seemingly immediate display of its associated symbol, or by some other appropriate display change.	
COMMENT: This recommendation is intended to ensure efficient operation in routine, repetitive data entry tasks. Longer delays may be tolerable in special circumstances, perhaps to reduce variability in computer response, or perhaps in cases where data entry comprises a relatively small portion of the user's task.	
COMMENT: Note that this guideline refers to acknowledgment, rather than final processing of entries which may be deferred pending an explicit ENTER action.	
SEE ALSO: 3.0-18, 3.0-19.	

+Opt

Figure 2. Detailed Individual Guideline

Hypertext Tool for Selecting Design Guidelines

REFERENCES

- Gould, J. D. and Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM*. 28(3), 300-311.
- Gould, J. D., Alfaro, L., Finn, R., Haupt, B., and Minuto, A. (1987). Reading from a CRT can be as fast as reading from paper. *Human Factors*, 29(5), 497-517.
- Hansen, W. J. and Haas, C. (1988). Reading and writing with computers: A framework for explaining differences in performance. *Communications of the ACM*. 1080-1089.
- Mosier, J. N. (1988). *Applying guidelines in rule-based user interface design*. (Unpublished report). The MITRE Corporation.
- Normore, L. F. and Tijerina, L. (1986). Evaluation of guidelines for designing user interface software. In *Proceedings of the Human Factors Society 30th Annual Meeting*. Dayton, OH. 1363-1365.
- Perlman, G. (1989b). System design in evaluation with hypertext checklists. In *IEEE Conference on Systems, Man and Cybernetics*, Cambridge, MA.
- Reaux, R. A. and Williges, R. C. (1988). Effects of abstraction and presentation media on usability of user-system interface guidelines. In *Proceedings of the Human Factors Society - 32nd annual meeting*. Anaheim, CA. 330-334.
- Salton, G. (1975). *Dynamic information and library processing*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Smith, S. L. (1986). Standards versus guidelines for designing user interface software. *Behaviour and Information Technology*, 5(1), 47-61.
- Smith, S. L. and Mosier, J. N. (1984). *A design evaluation checklist for user-system interface software*. (Report No. ESD-TR-84-358). The MITRE Corporation.
- Smith, S. L. and Mosier, J. N. (1986). *Guidelines for designing user interface software* (Report No. ESD-TR-86-278). The MITRE Corporation.
- Thovtrup, H. and Neilsen, L. (1991). Assessing the usability of a user interface standard. In *Human Factors in Computing Systems: CHI '91*. New Orleans, LA. 335-341.
- Tijerina, L. (1986). Design guidelines and the human factors of interface design. In *Proceedings of the Human Factors Society 30th Annual Meeting*. Dayton, Ohio. 1358-1361.
- Woods, D. D. and Eastman, M. C. (1989). Integrating principles for human-computer interaction into the design process: Heterarchically organized HCI principles. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. 29-34.

END

DATE
FILMED

2 / 9 / 93