BEHAVIOUR & INFORMATION TECHNOLOGY, MAY–JUNE 2003, VOL. 22, NO. 3, 155–163

Taylor & Francis
Taylor & Francis Group

The effect of language inconsistency on performance and satisfaction in using the Web: results from three experiments

A. ANT OZOK* and GAVRIEL SALVENDY†‡

*Department of Information Systems, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD, 21250, USA, e-mail: ozok@umbc.edu

†School of Industrial Engineering, 1287 Grissom Hall, Purdue University, West Lafayette, IN, 47907-1287, USA, e-mail: salvendy@ecn.purdue.edu

‡Department of Industrial Engineering, Tsinghua University, Beijing, PR China

Abstract. The objective of this study was to investigate the effects of linguistic inconsistency on performance and satisfaction on the Web. Three experiments were conducted, and the Interface Consistency Testing Questionnaire (Ozok and Salvendy 2001) was used in all three experiments to measure consistency levels. The three experiments using a total of 100 subjects evaluated the effects of direct manipulation, text-based, and hybrid interfaces with linguistic consistency and inconsistency on performance and satisfaction of the Web user. The experimental results indicated that both direct manipulation and text-based interfaces with linguistic consistency reduced performance time and error rates in comparison to inconsistent linguistics interfaces. Specifically in Reading Comprehension Tasks, subjects committed significantly fewer comprehension errors in consistent language interfaces than in inconsistent language interfaces.

1. Introduction

Consistency of interface design has been the concern of many researchers for more than a decade. Previous literature indicated the positive effect on performance of structural interface consistency of computer interfaces, however, the consistency of interface language and information representation have not been investigated in the previous literature. Based on the aforementioned issues, the main objective of this study can be identified as determining the impact of language consistency of interfaces on the performance and satisfaction of the World-Wide-Web user.

2. Background literature

Features of consistency include physical and communicational consistencies of the layout. *Physical consistency* deals with the physical features of interfaces such as colour and size whereas *communicational consistency* is concerned with the low-level interaction between the user and the interface, such as scrolling the screen with a mouse. Also, understanding the consistency of language is essential in designing interfaces that present information compatible with the user's understanding and perception.

Shneiderman's (1997) first rule for design is: 'Strive for consistency.' Myers and Fisk (1987), on the other hand, concluded that training of consistent task elements helps ease the training. Similarly, De Rosis *et al.* (1998) indicated that if the interface is not consistent, the field user will have difficulty in associating the external behaviour of the system with its internal functioning. Additionally, Adamson (1997) concluded that physical inconsistencies in an interface cause a significant loss of performance and user satisfaction.

Independent from computer interfaces, the effect of language and text attributes on the performance of people dealing with text has been the concern of many linguists (such as Kintsch and Van Dijk 1983). The primary goal of most of the linguistic researchers was to determine the elements that make the language more commonly and more quickly understandable. Linguistic consistency can therefore be understood in terms of understandability, clarity, and eligibility, and a well-formed interface language can be perceived as consistent.

Additionally, Kintsch (1998) made the distinction between the categories of interface text, with large emphasis on the microstructure and macrostructure.

Behaviour & Information Technology
ISSN 0144-929X print/ISSN 1362-3001 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0144929031000107081

Based on the definitions by Kintsch (1998) and Kintsch and Van Dijk (1983), seven different categories of interface text have been defined: concept, microstructure, macrostructure, propositions, general text features, lexical categories, and topic. Each of these categories is applicable and definable in a text-based interface, and consistency of each of these elements can be viewed as a performance improving factor for interface users. The detailed definitions of linguistic consistency are presented in the next chapter.

3. Definitions of linguistic consistency

The seven definitions of linguistic consistency have not been explicitly defined in the previous literature in the relation of interface design. Hence, although the definitions of consistency are presented briefly in Table 1, it was deemed essential to define those seven consistency definitions (consistency of concept, microstructure, macrostructure, propositions, general text features, lexical categories, and topic) in more detail.

3.1. *Consistency of concept*

Kintsch (1998) indicates that the concept of a text is a network of connections, a knowledge network, among meanings of individual parts (sentences and words) that build the text. Also according to Kintsch, although concepts do not have a fixed and permanent meaning, each time a concept is used, its meaning is constructed in working memory by activating a certain subset of the propositions in the neighbourhood of a concept node, and goals, prior experience, emotional state and situational and semantic context all influence which nodes are activated and hence what the meaning of the concept will be on that particular occasion. Hence, the consistency of concept can be defined as the meaning of this text concept being consistently understood among different users, leaving little room for argument of the meaning of the network of knowledge built by individual propositions (the concept of propositions is explained later as a sub-section in this chapter). It should be noted that although the idea of concept consistency might seem to be fairly broad, its main argument is for the concept of a text of an interface being understood consistently among different users.

3.2. *Consistency of microstructure*

According to Kintsch (1998), the microstructure is the local structure of the text, a sentence-by-sentence information, as supplemented by and integrated with long-term memory information. In the comprehension process, chunks of information presented by sentences are coupled with their meanings by the users. Hence the consistency of microstructure can be defined of the consistency of this coupling of sentence-by sentence representation of information with the correct and consistent understanding by all users. The meanings of the individual sentences are therefore to be consistently understood by the different users.

3.3. *Consistency of macrostructure*

Kintsch and Van Dijk (1983) indicated that the macrostructure is a hierarchically ordered set of propositions representing the global structure of text derived from the microstructure. In other words, macrostructure is a broader look of the meaning of information presented by the overall text. However, it is indicated by Kintsch that this overall meaning is in turn built by the microstructure of the text. Hence, the consistency of macrostructure can be defined as the consistency of the connection between the individual parts (sentences) of the text and the overall meaning that can be derived from the text. If the connections between the overall text and its individual parts (sentences) are perceived by all users the same way, then it can be said that the consistency of macrostructure is present in the text.

3.4. *Consistency of propositions*

In the simplest sense, a propositional representation is the mental representation of text items by the user. The propositions primarily reflect semantic relations between what the text is intending to convey to the user and what the user can comprehend. Hence, if a proposition is presented, its consistency with how people understand, remember, and think in language-related terms is essential to its comprehensibility.

3.5. *Consistency of general text features*

Since the concept of linguistic consistency is defined as an element of interface design that is an integral part in increasing the comprehensibility of interface text, it is deemed essential that the essential elements of comprehensibility be integrated to the concept of consistency. Hence, avoidance of ambiguity, cumbersomeness, confusion or contradiction of text elements is seen as essential within the context of consistency. According to Kintsch, these general factors of text are believed to play an

Table 1. Descriptive statistics and differences in means values for the performance variables in three experiments.

	Experiment 1				Experiment 2		Experiment 3	
Group	Direct man. consistent ($N = 10$)	Direct man. inconsistent ($N = 10$)	Text-based consistent ($N = 10$)	Text-based inconsistent ($N = 10$)	Consistent ($N = 20$)	Inconsistent ($N = 20$)	Consistent ($N = 10$)	Inconsistent ($N = 10$)
Descriptive statistics*	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)
Performance time (sec.)								
Overall	2543.6 (584.90)	2725.3 (685.44)	2127.5 (555.90)	2738.1 (645.64)*	2254.4 (418.31)	2750.8 (692.79)*	2228.1 (418.85)	2852.2 (427.77)*
Reading	1117.7 (226.82)	1353.6 (292.10)	979.2 (390.71)	1429.0 (347.91)*	1050.2 (208.77)	1325.5 (459.37)*	948.7 (256.10)	1303.5 (311.01)*
Word search	1032.7 (364.94)	980.1 (347.45)	803.0 (205.55)	912.8 (342.41)	872.6 (289.20)	1022.0 (275.43)	933.7 (276.83)	1170.0 (270.83)
Typing	159.0 (50.70)	158.1 (47.59)	112.4 (28.18)	135.0 (43.53)	135.1 (37.18)	166.2 (44.06)*	158.4 (39.17)	163.5 (44.82)
Form filling	234.2 (44.73)	238.9 (51.44)	232.9 (66.36)	261.3 (66.72)	196.6 (32.21)	237.2 (61.64)*	187.3 (51.19)	229.4 (58.32)*
Confusion								
Errors overall	2.7 (1.89)	3.1 (2.85)	2.4 (3.53)	1.6 (1.35)	1.8 (2.10)	4.27 (3.82)*	2.7 (4.64)	6.8 (3.05)*
Reading***	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Word search	1.8 (1.55)	2.4 (2.67)	1.7 (2.45)	0.8 (0.79)	1.3 (1.45)	3.1 (3.46)*	1.9 (3.51)	5.6 (2.59)*
Typing	0.8 (1.14)	0.7 (0.82)	0.5 (1.58)	0.7 (0.95)	0.5 (0.83)	0.7 (1.03)	0.8 (1.32)	1.1 (1.10)
Form filling	0.1 (0.32)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Comprehension								
Errors overall	2.6 (0.84)	5.8 (3.19)*	2.0 (1.76)	3.7 (1.57)*	2.0 (1.85)	4.9 (3.42)*	0.9 (1.20)	3.1 (4.04)
Reading	2.4 (0.70)	4.1 (2.28)*	1.4 (1.35)	2.3 (1.42)	1.7 (1.75)	3.4 (2.30)*	0.4 (0.70)	1.4 (1.51)
Word search	0.2 (0.42)	1.6 (2.17)	0.6 (0.97)	0.9 (0.99)	0.3 (0.55)	1.4 (1.87)*	0.4 (0.97)	1.7 (3.16)
Typing	0 (0)	0 (0)	0.1 (0.32)	0.1 (0.32)	0 (0)	0 (0)	0 (0)	0 (0)
Form filling	0 (0)	0.1 (0.32)	0 (0)	0.4 (1.26)	0 (0)	0.1 (0.45)	0 (0)	0 (0)
Combined errors								
Overall	7.9 (2.69)	14.7 (6.18)*	6.4 (3.72)	9.0 (3.40)	5.7 (4.72)	13.5 (8.41)*	4.5 (6.11)	13.0 (7.09)*
Reading	4.8 (1.40)	8.2 (4.57)*	3.0 (2.54)	4.7 (2.83)	3.4 (3.50)	6.8 (4.61)*	0.8 (1.40)	2.7 (2.98)
Word search	2.2 (2.04)	5.6 (4.43)	2.7 (2.63)	2.6 (2.50)	1.8 (1.96)	5.8 (5.13)*	2.7 (5.23)	9 (5.46)*
Typing	0.8 (1.14)	0.7 (0.82)	0.7 (1.64)	0.9 (1.20)	0.5 (0.83)	0.7 (1.03)	1.0 (1.49)	1.1 (1.10)
Form filling	0.1 (0.32)	0.2 (0.63)	0	0.08 (2.53)	0 (0)	0.2 (0.89)	N/A	N/A

*: The numbers marked (*) are significantly different from the numbers on their left at 0.05 level. **: The numbers in brackets are standard deviation values. ***: Confusion Errors are not applicable to Task 1 because Task 1 only involves multiple-choice questions where subjects are not required to type and the multiple choices are located far enough from each other to prevent the subjects from mis-clicking.

essential role in building comprehensible text and therefore are included within the definitions of consistency.

3.6. *Consistency of lexical categories*

All of the words within a text can be classified into the one of the lexical categories of nouns, verbs, adjectives, adverbs and adpositions. The consistency of lexical categories indicates the words being consistently and correctly placed across the text, not violating the correct positioning and being consistent with the knowledge of the user regarding the use of these lexical elements.

3.7. *Consistency of topic*

The topic of a text is usually indicated by its title. The consistency of the topic indicates a strong relation mainly between the individual parts (sentences or paragraphs) of the text and its title as well as a relation between those individual parts themselves. If the explanations in the paragraphs of the text are deviating from the title of the text, a focused explanation as well as topic consistency would be lacking.

3.8. *Distinction between internal vs. external consistencies*

Within the concept of consistency, it needs to be mentioned that the scope of linguistic consistency covers the *internal* consistency issues in the human-computer environment, meaning the consistency of interface elements with each other within an interface. *External* consistency of interfaces, where the consistency of different screen elements between different interfaces is in focus, has not been taken into consideration within the scope of linguistic consistency. More information regarding this distinction can be found in the study conducted by Ozok and Salvendy (2000).

4. Experimentation on the effects of language consistency

4.1. *Overview*

Four experiments were conducted in order to determine the effect of interface language consistency on the performance and satisfaction of the users on the World Wide Web. Consistency levels of interfaces were measured by two different questionnaires in the first three experiments. These questionnaires were the Interface Language Consistency Testing Questionnaire (ILCTQ) and the Purdue Consistency Testing Questionnaire (PCTQ; Ozok and Salvendy 2000). These two questionnaires were later combined and a single explicit tool for measuring the consistency levels of interfaces was developed, called the Interface Consistency Testing Questionnaire (ICTQ). All three questionnaires used 7-point Likert scales. The first three experiments were conducted on English major students because of the high number of linguistics-related terms used in the questionnaires. More details regarding the experimental design, the measurement tool for consistency, and screen outputs distinguishing between the consistent and inconsistent designs can be found in the study by Ozok and Salvendy 2001.

4.2. *Experiment 1: Measuring the effect of language consistency and its relationship to interface orientation*

4.2.1. *Methodology*: It was hypothesized that interfaces with consistent language would result in significantly decreased error rates and performance times, and that there would be a significant interaction between the orientation and language consistency of interfaces. Therefore, the performances of the users of consistent and inconsistent interfaces were measured. The interfaces were distinguished in terms of interface consistency (consistent vs. inconsistent interfaces) and interface orientation (direct manipulation vs. text-based interfaces). Satisfaction levels of the users were also measured by presenting a modified version of the Hackman-Oldham Job Satisfaction Questionnaire (Hackman and Oldham 1980) consisting of sixteen questions.

4.2.1.1. *Independent and dependent variables*: The independent variables were interface orientation and interface language consistency. Interface orientation is defined as being either direct manipulation or text-based. Interface language consistency was measured by the Interface Language Consistency Testing Questionnaire (ILCTQ). It is defined as the overall score obtained from this questionnaire. The dependent variables were performance time and number of errors. Performance time for each task was measured by the PERL computer programme. The performance time was recorded to the nearest second for each stage of each task separately. Number of errors was measured by collecting the data from the computer and then manually finding the errors the subjects made. Different types of errors had different weights during the experiment. A typing error or clicking accidentally on the wrong item on the screen are called a Confusion Error. Errors involving comprehension and interpretation are called Comprehension Error, are considered more serious and are given twice

the weight of Confusion Errors in the data analysis. For the purpose of obtaining the data for later combining the two questionnaires, the Purdue Consistency Testing Questionnaire (PCTQ; Ozok and Salvendy 2000) was also presented to the users. Both questionnaires, the ILCTQ and the PCTQ, made a significant distinction between consistently and inconsistently designed interfaces, and had internal reliabilities of 0.83 and 0.78, respectively, and interrater reliabilities of 0.73 and 0.74, respectively.

4.2.1.2. *Subjects*: Students from Department of English at Purdue University with at least 1 year of experience on the World Wide Web and who were using the Web frequently (at least three times a week) were recruited as participants in the experiment. Of the 40 participants (age mean = 30.1, Std. Dev. = 8.46) in the experiment, 20 were female and 20 were male.

4.2.1.3. *Experimental design*: The experimental design was a two-factor, between-subjects design. There were four groups in the experiment: Group 1 had direct manipulation interfaces with consistent language; Group 2 had direct manipulation interfaces with inconsistent language; Group 3 had text-based interfaces with consistent language; and Group 4 had text-based interfaces with inconsistent language. Each group had 10 subjects. Subjects were randomly assigned to one of the four groups. In order to avoid variations among the groups because of gender, male and female subjects were balanced in each group.

4.2.1.4. *Task*: There were four stages in the task. The first stage was a reading comprehension task. Subjects were presented five screens containing substantial amounts of text. After reading the text on each interface, subjects were required to click on a hyperlink that transferred them to a page in which they were asked five questions regarding the text. The second stage was a word search task. There were six screens in this task, and subjects were to search for one word in each task. The third stage was a typing task in which subjects were provided a one-paragraph text on one screen and were required to type the same paragraph into the specified area at the bottom of the screen. The fourth and final stage consisted of a form-filling task. There were eight different screens for this particular task. In this task, subjects were required to single-click on specified screen buttons, radio buttons, or checkboxes on the screen with the mouse. They also opened up drop-down combo boxes and changed them by mouse. All of the taskbars and menus were deactivated during the entire experimentation, and the subjects were instructed not to open or close windows during the experiment.

In the interfaces with direct manipulation attributes, the user was frequently required to do operations with the mouse, and objects such as hyperlinks and split-screens were frequently used. The text was present for the tasks, but the interfaces were manipulated by different actions of the user. In interfaces with text-based attributes, interface elements besides text were used minimally. Mouse operations were mostly conducted on text or hypertext elements of the interface. Text was used much more frequently than other interface elements in those interfaces. Three direct manipulation features were integrated to direct manipulation interfaces: Scrolling, split-screen, and within-page hyperlinks that allow the user to jump to the beginning of a chapter on the page.

These elements of consistency discussed in the previous chapter were used in manipulating the Web pages used in the task and in producing consistent and inconsistent versions. Examples of consistent and inconsistent interfaces can be found in the study by Ozok and Salvendy 2001.

For the distinction between interfaces with consistent and inconsistent language, elements of interface language consistency were manipulated. In the consistent version of interfaces, the content of the text was manipulated to be consistent with the linguistic knowledge of the user; the nouns, verbs, adjectives, adverbs, and adpositions were intended to be consistent with the knowledge of the user and located consistently throughout the text; information was presented in a focused way; the sentences used consistent tenses; the wording was consistent with the topic of the text; a hierarchy of subtopics was promised; confusion, ambiguity, contradiction, and cumbersomeness were avoided; clarity was observed; and the overall text layout tried to be consistent within an interface and between different interfaces. About 30 items of consistency (or inconsistency in the inconsistent versions) were integrated to each page.

In the interfaces with inconsistent text elements, all of these specifications were avoided. In the consistent text version of each interface, the abbreviations of the names (such as MS for Master of Science) are presented in parentheses, whereas the abbreviations are used without any explanations in the inconsistent versions (as part of consistency of general text features). Additionally, several adjectives and words belonging to other lexical categories are used inconsistently in the inconsistent version (such as the word 'scholastical' in the inconsistent interface).

4.2.2. *Results and discussion*

4.2.2.1. *Descriptive statistics and analysis of variance for performance variables*: Descriptive statistics values for Experiment 1 were calculated for all of the performance

variables in their overall as well as decomposed (according to tasks) forms and are presented in Table 1. In order to prevent any problems with Type I error, a Bonferroni correction procedure was applied for confirming the differences in means of performance variables detected by the analysis of variance.

The results indicate that Overall Time significantly differs between consistent and inconsistent text-based interfaces, but not in direct manipulation interfaces. The same can be said for Confusion Errors for the Overall Task. In the text-based consistent interface language group, subjects performed significantly better in terms of Task Completion Time than in the text-based inconsistent language group. However, when the orientation was direct manipulation, meaning more interactive with less emphasis on the text, then the inconsistency of language interface had no significant effect on the Overall Performance Time in comparison to the consistent one. This indicates that increasing the emphasis on the text in interface orientation resulted in Task Completion Times being significantly reduced when consistent language is presented, but had no effect in Word Search and On-Line Form-Filling tasks. Part of this result can be explained by the Form-Filling tasks being very easy in nature. Also, it should be taken into consideration that variance in typing speed is not a result of interface consistency but rather individual ability, and the sole goal of the Typing Task was to measure this.

The errors were significantly different in direct manipulation and text-based interfaces between consistent and inconsistent groups for Overall Comprehension Errors. In Comprehension Errors in Reading Comprehension Task and Confusion and Comprehension Errors Combined in the Overall Task and Reading Comprehension Task, the differences were significant between consistent and inconsistent interfaces in direct manipulation interfaces, but not in text-based interfaces. The results indicate a fair consistency in error variables between direct manipulation and text-based interfaces. However, reading comprehension was significantly improved in terms of comprehension and overall errors in consistent groups with direct manipulation interfaces, but not with text-based interfaces.

In order to measure the availability of the interaction variable between interface orientation (direct manipulation vs. text-based) and interface consistency (consistent vs. inconsistent), two-way analysis of variance was conducted on the overall performance variables. This particular analysis did not indicate a significant interaction value in the Overall Time and Overall Error variables.

Overall, it can be concluded that in direct manipulation interfaces, significant differences were detected in overall comprehension errors due to the comprehension errors committed in the Reading Comprehension Task, whereas in almost all text-based interfaces, Overall Task Completion Time was significantly reduced with the introduction of interface language consistency. This leads to the general conclusion for Experiment 1 that interface language consistency results in significantly better Task Completion Times in most text-based interfaces because of increased concentration on text-related issues in these interfaces. Also, errors are significantly reduced for Reading Comprehension Tasks in direct manipulation interfaces when interface language consistency is introduced, which may be because of the more interactive nature of the interfaces. No significant interactions were detected between interface orientation and interface consistency values.

There was no statistically significant change in user satisfaction as measured by the 16-item structured General Satisfaction Questionnaire with an internal reliability value of 0.83 and an interrater reliability value of 0.81. This particular finding is in accordance with previous findings by Ozok and Salvendy (2000) who concluded that the introduction of interface consistency did not result in user satisfaction, mainly because of potential boredom experienced by the users.

4.3. *Experiment 2: Determining the correlations between interface consistency elements*

4.3.1. *Methodology*: Experiment 2 tested the effect of the interface language consistency on the performance and the satisfaction of the user, without the distinction between direct manipulation and text-based interfaces. Since an interaction was expected between interface consistency and interface orientation, Experiment 2 had the goal of measuring the effect of interface language consistency without the presence of the interface orientation variable. The interfaces were distinguished between consistent and inconsistent languages, with 40 subjects in the entire experiment and 20 subjects in each of the consistent and inconsistent groups. The methodology was very similar to the first two experiments. The study of the relationship between interface consistency and performance without the involvement of interface orientation and information representation was the main reason in conducting Experiment 2. Also, Experiment 2 and the subsequent Experiment 3 played a significant role in conducting the factor analysis in order to finalize the consistency measurement tool. Details regarding this tool can be found in a separate study by Ozok and Salvendy (2001).

4.3.1.1. *Independent and dependent variables*: The independent variable was interface language consistency. It was measured by the Interface Language Consistency Testing Questionnaire (ILCTQ). The dependent variables were performance time, number of errors, and satisfaction scores. The PCTQ was again presented to the subjects for combination purposes. The reliability and validity values of both questionnaires were acceptable. The Cronbach's Alpha values for internal reliability were 0.81 for the ILCTQ and 0.85 for the PCTQ. The interrater reliability values were 0.77 for the ILCTQ and 0.74 for the PCTQ.

4.3.1.2. *Subjects*: Each group had 20 subjects, with a total of 40 subjects. The same specifications and qualifications as in Experiment 1 were required from the subjects in Experiment 2. Of the 40 participants (age mean = 27.4, Std. Dev. = 5.86), 24 subjects were female and 16 were male.

4.3.1.3. *Experimental design*: The experimental design was a single-factor, between-subjects design. There were two groups in the experiment: Group 1 had consistent interface language, and Group 2 had inconsistent interface language. Subjects were randomly assigned to one of the two groups. The distinction between consistent and inconsistent language interfaces was made through the same procedure of interface manipulations described in Experiment 1. Again, subjects were balanced in each group in terms of gender.

4.3.1.4. *Task*: The task sequence in the experiment was the same as in Experiment 1. There was no distinction between text-based and direct manipulation interfaces in this experiment. The interfaces only differed in consistency of language, and since the distinction in interface orientation (direct manipulation vs. text-based) was not made in the experiment, the interfaces were text-based.

4.3.2. *Results and discussion*

4.3.2.1. *Descriptive statistics and analysis of variance for performance variables*: The descriptive statistics and analysis of variance for performance variables in Experiment 2 were analysed and are presented on Table 1. A Bonferroni correction procedure was applied again.

The Task Completion Time was significantly different between consistent and inconsistent interfaces in almost all of the tasks, with the exception of the Word Search Task. The Task Completion Time was significantly shorter in consistent interfaces, in accordance with the findings from the two previous experiments as well as previous studies such as the one conducted by Adamson and Wallace (1997). The results lead to the conclusion that in most tasks, interface language consistency results in reduced task completion times of Web users. However, in the Word Search Task, significant differences in times were not detected (17.12%, $F_{1,38} = 2.80$, $p = 0.1025$). This leads to the conclusion that solely changing the interface language consistency results in no performance time increase in searching. It should also be noted that, as was the case in Experiment 1, Typing Task resulted in significant differences in Task Completion Time, although the interfaces were not contrasted in this particular task. This finding again indicates between-subject skill differences. Overall, the conclusion could be made that interface language consistency results in a significant decrease in Task Completion Time of the computer interface user.

For the majority of the tasks, the number of errors was significantly higher in the inconsistent group, except for the Confusion Errors in the Overall Task. This particular finding can be traced back to the fact of the variance being very high and the data points being few. It can be concluded that increasing the language consistency level on Web-based computer interfaces results in decreased number of errors. Based on the findings, it was concluded that the overall errors and time variables were significantly improved when the language consistency levels of interfaces were increased.

Similarly to the first two experiments, the differences between the satisfaction scores of the consistent and inconsistent interface users were measured, and no significant differences between the two groups were detected.

4.4. *Experiment 3*

4.4.1. *Methodology*: After the first three experiments, the two questionnaires, the PCTQ and the ILCTQ, were combined into one structured questionnaire. The total number of questions in this new questionnaire was reduced from 125 to 94 through a factor analysis. This newly developed questionnaire was called the Interface Consistency Testing Questionnaire (ICTQ), and it had 94 questions. The reason behind conducting Experiment 3 was to measure the construct validity and reliability of the ICTQ and to understand whether results obtained from English major participants were applicable to a more general population. For this latter reason, students from the Schools of Engineering at Purdue University were used in the experiment. The experimental design was a duplicate of Experiment 2.

4.4.1.1. *Independent and dependent variables*: The independent variable was interface language consistency. It was measured by the newly developed Interface

Consistency Testing Questionnaire (ICTQ; overall and in the four categories, Consistency of Sentences and General Text, General Language Features, Words, and Physical Features). The dependent variables were performance time and number of errors. The 94-question, 9-factor questionnaire had an internal reliability of 0.81 and an interrater reliability of 0.75.

4.4.1.2. *Subjects*: There were 20 subjects (age mean = 26.6, Std. Dev. = 3.03) in the experiment. All of the subjects were taken from the different departments belonging to the Schools of Industrial Engineering at Purdue University, and all were graduate students. All of the subjects were required to have the same specifications as in the first two experiments.

4.4.1.3. *Experimental design*: The experimental design was again a single-factor, between-subjects design. There were two groups in the experiment: Group 1 had consistent interface language, and Group 2 had inconsistent interface language. Subjects were again randomly assigned to one of the two groups, and male and female subjects were balanced in each group.

4.4.1.4. *Procedure*: The task sequence in Experiment 3 was the same as in the previous two experiments. There was no distinction between text-based and direct manipulation interfaces. The interfaces only differed in consistency of language.

4.4.2. *Results and discussion*

4.4.2.1. *Descriptive statistics, comparison of means, and analysis of variance results for performance variables*: The performance variables were compared between the consistent and inconsistent groups and are presented on Table 1. Again, a Bonferroni correction procedure was applied in order to prevent any problems with Type I error.

The results presented are mostly consistent with the results obtained from the previous experiments, especially Experiment 2. The Overall Task Completion Time was significantly shorter in the consistent group than in the inconsistent group (28.01%, $F_{1,18} = 10.87$, $p = 0.0040$). It was concluded that inconsistent interfaces did not result in increased task completion time for Word Search Tasks. The finding of no differences in time for the Typing Task was expected because there were no differences in consistencies in this task. And consistently with Experiment 2, there was a significant difference in Task Completion Time in the Form-Filling Task (22.48%, $F_{1,18} = 4.29$, $p = 0.0529$).

Additionally, there was a significant difference in Overall Confusion Errors (151.85%, $F_{1,18} = 5.45$, $p = 0.0314$). The findings regarding the Confusion Errors are entirely consistent with Experiment 2 in terms of significant differences. In Experiment 3, Comprehension Errors significantly differed in the Word Search Task (194.74%). Contrary to Experiment 2, in Experiment 3, there were no significant differences in Comprehension Errors. This leads to the conclusion that for Engineering students, comprehension of what they read was equally good when inconsistency of language was introduced. Also, combined errors differed significantly for the Overall Task (188.89%, $F_{1,18} = 15.20$, $p = 0.0011$) and for Word Search Task (233.33%, $F_{1,18} = 8.42$, $p = 0.0037$). As expected, there were no significant differences for Typing Task, and because of the lack of data points, no analysis of variance was performed for error variables in the Form-Filling Task. Overall, it was concluded that the performance results were consistent with the results from Experiment 2 on English Department students.

As indicated in the previous chapters, the results from the previous experiments indicated no significant differences in satisfaction levels when consistency was introduced.

5. Conclusions

Authors such as Shneiderman (1997) mentioned consistency as being part of usability, but its significance had not been extensively investigated previously. Also, although several researchers, such as Kintsch (1998), partially investigated consistency of the English language, the concept of language and its consistency had not been integrated into the area of human-computer interaction. This research attempted to realize this integration.

In the current study, the user task completion time, number of errors, and satisfaction were taken as a function of interface language consistency, interface orientation (direct manipulation vs. text-based), and interface tasks. The results of the four experiments regarding the effects of the inconsistency of language presentation on performance and satisfaction indicated the following:

- The consistency of interface language is a significant element of interface design.
- If consistency rules regarding language are obeyed in the overall design, the task completion times and number of errors of users of computer interfaces on the World Wide Web will be significantly improved.
- Consistency of interface language improves user performance especially in tasks where information is retrieved from the interface through reading,

and when the user is looking for specific information (in terms of a word) on the interface.
- Consistency of language resulted in reduced error rates and task completion times, but not higher satisfaction, which can be traced back to the boredom factor that arises with increased consistency.
- Interface language consistency is a significant element of overall consistency of computer interfaces on the World Wide Web.
- The newly developed Interface Consistency Testing Questionnaire was successful in distinguishing between consistent and inconsistent language interfaces, and has high internal (0.81) and interrater (0.75) reliabilities.

References

Adamson, P. J. 1997, *A comparison between consistent and Inconsistent graphical user interfaces*. Unpublished report, University of North Florida, Department of Computer and Information Sciences, Jacksonville, Florida, USA.

De Rosis, F., Pizzutilo, S. and De Carolis, B. 1998, Formal description and evaluation of user-adapted interfaces. *International Journal of Human Computer Studies*, **49**, 95–120.

Hackman, R. J. and Oldham, G. R. 1980, Development of job diagnostic survey. *Journal of Applied Psychology*, **60**(2), 159–170.

Kintsch, W. 1998, *Comprehension: a paradigm for cognition* (Cambridge, UK: Cambridge University Press).

Kintsch, W. and Van Dijk, T. A. 1983, *Strategies of discourse comprehension* (New York, USA: Academic Press).

Myers, G. L. and Fisk, A. D. 1987, Training consistent task components: application of automatic and controlled processing theory to industrial task training. *Human Factors*, **29**(3), 255–268.

Ozok, A. A. and Salvendy, G. 2000, Measuring consistency of Web page design and its effects on performance and satisfaction. *Ergonomics*, **43**(4), 443–460.

Ozok, A. A. and Salvendy, G. 2001, How consistent is your web design? *Behavior and Information Technology*, **20**(6), 433–447.

Shneiderman, B. 1997, *Designing the user interface: strategies for effective human-computer interaction*, 3rd edn (New York, USA: Addison Wesley Longman).