

Cross-Use: Cross-Cultural Usability User Evaluation- In-Context

Jasem M. Alostath and Abdulwahed Moh Khalfan

Faculty of Computing & Information Systems Department, Collage of Business Studies,
PAAET, Kuwait

P.O.Box 23167, Safat, 13092, Kuwait

m_jasem@yahoo.com, dr.khalfan@yahoo.com

Abstract. This paper introduces the Cross-Use experiment, which aims to evaluate the mapping between website design elements and cultural attributes using a user-in-context evaluation approach. This is done by developing three UI designs, and applying them to 63 local participants from the case study cultures (UK, Egypt, and Kuwait). The experiment was conducted using the developed prototypes was able to classify cultures differently, and highlighted those design markers that affects cultural differences in the design of e-banking websites. This is based on user preferences and usability.

Keywords: Culture, Usability, User preferences, e-banking, user-in-context evaluation.

1 Introduction

Many cross-cultural design evaluations use existing websites designs in identifying cultural design differences. However, these design evaluations are not supported with a cultural model, or adopts cultural models that are not design oriented in interpreting design based on culture [5, 6, 7, 8]. In our research of Culture-Centred Design (CCD) we have conducted design evaluations based on the identified subjective cultural attributes (CA) that characterize similarities and differences within and between user groups of different nationality of the cultural model that were developed based on HCI design [9, 13]. The most important advantage of this new approach is that the results of the analysis provide the designer with sufficient information to generate new websites that are more sensitive to culture and genre variability. However, the designs generated are not guaranteed to be optimal. This is because: (1) the existing websites that form the basis of the analysis may not have been well designed from the cultural point of view, (2) the claims from the cultural-design mapping from which designs are generated may be insufficient to determine a unique design decision, and (3) the design analysis that is undertaken does not provide any important information on design aspects such as usability [9]. Our solution to this problem is in the CCD methodology [9], which uses the design analysis results to develop a number of possible prototype websites that will be culturally adapted to some degree. Then a rigorous user testing approach is used to decide between the alternatives (further details about the CCD method see Alostath [9]).

2 Cross-Use: Method and Process

The experiment design involves three national cultures, using three user interfaces for simple and complex tasks (3*3*2 mixed design). The independent variables of the cultural factors were manipulated using three designs and are shown using the Latin Square design to counterbalance order effects [1]. The prototype used in this experiment was developed from scratch by the researchers based on the results of the design analysis. The three websites developed have one user interface design for each culture that maximizes the cultural and genre attributes appropriate for that culture. In addition, for each of the interfaces developed design alternative with content that is appropriate for each of the other cultures being tested is also included. This is done by exploiting the XML technology¹.

2.1 Variables and Participants

84 user variables are measured in this experiment. Fourteen variables are required to collect participants' demographic information. Of the remaining 70 variables, 58 are the users' subjective valuations of interface properties (e.g. text, images, and others) that are thought to have a cultural impact. The remaining 12 variables are used for evaluating each group of tasks (simple and complex tasks). Each group has six variables, of which four measure usability and two measure culture and trust compliance. These six variables are repeated for each task group. These 12 questions are aimed at building a usability factor that can be used to determine: (1) at the high level, the most usable design for each of the studied cultures, and (2) at the lower level, the design markers² (DMs) that improve usability from the 58 DMs. The experiments were conducted with 21 participants from each culture (Kuwait, UK, and Egypt). Participants were selected based on their ability to use the computer, internet, speak English and were given financial incentive.

2.2 Procedure and Materials

The Cross-Use experiment procedure consists of seven stages as shown in Figure 1. In the first stage, participants were informed about the three experimental sessions, objectives and procedure, and were required to sign the consent form. This is followed by the second stage, where each participant receives two 3-digit personal account codes and a password that allows them to run the experiment process and perform the online transactions required.

In the third stage, a questionnaire of 28 questions is administered; each question included one or more images of a DM relevant to one of the design claims being investigated. The aim is to obtain an initial understanding of the participants' expectations before interacting with the e-banking prototype. In the fourth stage (Task performance evaluation), the participant starts to perform six tasks, which are divided

¹ XML usually used to display different data across different UI platforms (e.g. Computer UI, mobile interface and others). Here, it is used to display different cultural data into HTML file, and this is based on users' culture.

² Design marker is a concrete design aspect and its existence is expected to have a cultural or genre, or other relations.

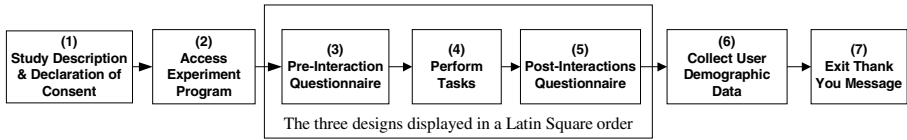


Fig. 1. Cross-Use experiment procedure

into two task groups (simple and complex tasks). Each group contains three tasks, the first three are for information inquiry and the other three are for performing transaction tasks. Upon completion of the three tasks, a comparison questionnaire is administered to rank the tasks. After each of the three tasks, participants answer the six design comparison questions, which compare the three designs in terms of usefulness, ease of use, frustration, satisfaction, culturally related issues and the most trustable design. The aim of this stage is to obtain the most usable design and what are the DMs that make a design usable for a particular culture. In the fifth stage, the participants were presented with several design layouts, and transactions processes necessary to explain the question, and were asked design-specific questions to rank several cultural design claims (30 questions presented in a forced-choice comparisons as well as 5-point Likert scale questions). The aim of this stage is to measure users' experience after their interaction with different interface designs and performing different types of tasks. The final stages are used to wrap-up the experiment by collecting participants demographic data and ending with a thank you message.

The experiment uses a Pentium Centrino 1.5 MHz laptop with 15" TFT screen, and regular mouse. The experiment was executed from the local web-server running on the same computer. In addition, a reasonable resolution (320 x 240 pixels) webcam was connected to the computer to record the participants' facial expressions using Morae™ tasks recording tool (see www.techsmith.com).

2.3 Objectives and Hypotheses

The objective of the Cross-Use experiment is to substantiate the cultural design claims [9, 12], which have been substantiated earlier in design evaluations approaches [9]. This experiment further substantiates these claims based on user-in-context evaluation, and aims to provide two types of results. These are related to the user preferences, and usability for the selected design, and design markers. User preferences refer to the results based on a comparison made by the user between two or more UIs or on specific aspects of those designs. In contrast, usability is assessed by performing real tasks, and then both objective (e.g. time to perform a task) and subjective (e.g. satisfaction with task) outcomes are measured. The results of users' preferences and usability are also useful in deciding whether the design preferences are a good indicator for usability. In order to test these objectives, several analysis methods were conducted, to examine the validity of the following hypotheses:

H1: When given a choice between a website designed for a different target culture and one designed for their own target culture, users will prefer the website designed for their own culture.

H2: Websites that have been designed for a particular target culture (e.g. Kuwait, or Egypt, or UK) using the developed cultural design claims will produce better usability results when tested by members of that particular target culture.

H3: Using Discriminant Analysis (DA), it is possible to identify specific or aggregated DMs that are the main contributors to the observed user preferences and usability improvement.

In this study, the DA and Chi-Square statistical analysis methods were used to analyse the questionnaire data, which involves a 189 observations -- 63 observations for 3 designs. The DA is used to show the most important or interpretive independent variables, which discriminate the dependent variable or affect it [11], while the Chi-square is used to determine whether the groupings of cases on one variable are related to the groupings of cases on another variable [2].

3 The Cross-Use Experiment

The aim of the Cross-Use experiment is to present the important DMs that were identified by users' preferences, and usability. This can be determined by two analyses, which are concerned with the ability of the developed user interface designs to classify the cultures differently, and the identification of those DMs that play a significant role in causing these differences. The key factors in this analysis are usability and preferences.

3.1 Cross-Cultural Design Preferences

Study hypothesis (H1) predicted that when creating designs that are in accordance with cultural design claims [9], these designs are able to generate culturally sensitive designs. The data collected from the experiment were used in this analysis to classify the three cultural groups of users according to their preferences for the identified cultural designs. DA was performed with national culture as the dependent variable, and the DMs as independent variables. The results of this analysis confirmed hypothesis H1 (see Figure 2 and Table 1). This indicates the ability of the website designs that adopted the cultural design claims to design for different cultures to capture users' different preferences. The DMs that cause the cultural preference differences among specific national cultures resulting from the above DA test are shown in Table 1.

3.2 Cross-Cultural Design Usability

In this section, an investigation of a good representative score for the cultural usability factor is conducted. Then, two types of analysis are performed. The first analysis uses a Chi-square test, and the second uses DA. The first analysis tells whether or not there is a relation between national cultures and design usability. The second analysis helps in classifying designs according to cultural usability and DMs, and identifying the DMs that are used to improve usability for each culture.

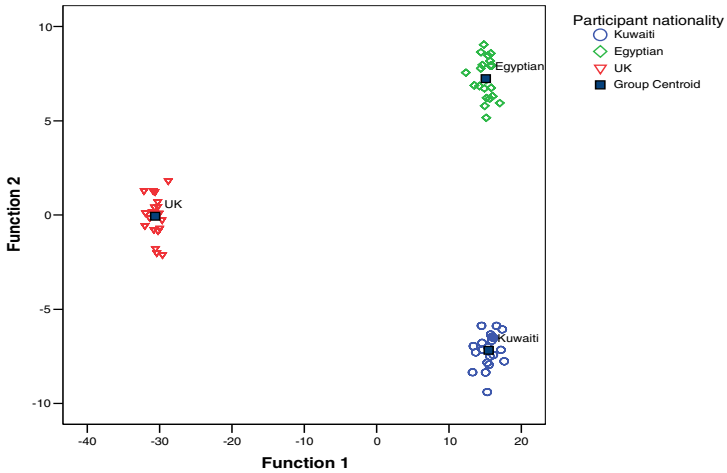


Fig. 2. Canonical Discriminant Functions plot: visualizing how the two functions discriminate between cultural groups by plotting the individual scores for the two functions

Table 1. Partial summary table for the user preferences DMs

CA	Claim	Design markers	KU	EG	UK	Related Question
		Relationship Metaphors				
R6, R7	C16 ³	Religious Metaphors (Design A)	M	M	L	B2a (*)
		National Metaphors (Design B)	M	H	M	B2b (*)
		Neutral Metaphors design (Design C)	H	H	H	B2c
		Navigation tools				
T4	C21	Drop-down Menu (complex navigation)	H	M	H	A1a (*)
		Tree-view (complex navigation)	L	M	L	A1b
		Sense of security				

Legend

CA is refer to the cultural attribute code identified in the HCI-cultural model [see 10]

- Low (L): <2.49; Medium (M)=2.50..3.49; High (H): >3.49

- (*) DM identified to be significant (p<.001) based on both the DA with Univariate ANOVA tests

- No sign indicates the DM was significant based on DA (p<.001) but not significant across cultures based on the Univariate ANOVA test (p<.001).

Culture and Usability Relation. The aim of this analysis is to identify the design differences affecting usability among the three cultures, based on the usability factor. Here, attempts are made to find if there are any relationships between national cultures and design usability. If there are any, then the DMs that are affecting usability across

³ Claim (C16): High racial tendency oriented cultures (relationship) are expected to show high use of religious and/or national symbols in the design more than low racial tendency oriented cultures, which tend to show neutral symbols.

these cultures are investigated. The study hypothesis (H2) predicts that when creating designs for cultures based on the cultural design claims and design investigation results (presented in [9]), such as design (A) for Kuwait, design (B) for Egypt, and design (C) for UK, such designs are expected to show better usability results by members of those particular cultures in their own cultural designs. Based on study hypothesis (H2), two issues need to be verified: the first issue is in determining whether a relation exists between culture and usability, which was verified using a Chi-Square test. Then, the second issue is determining the usability improvement that occurs frequently within the targeted cultural design, which was verified using a DA test.

As for the existence of a relation between the design usability (represented by the usability factor) and the national cultures, the following hypothesis was defined:

Hypothesis: There is a relation between national cultures and designs' usability (dependent)

A Chi-squared analysis shows that there is a significant relation between national culture and design usability ($\chi^2=19.08$, $df = 4$, $Sig. < .001$). In Figure 3, certain website designs are found to be more usable by certain national cultures is shown.

In validating hypothesis (H2), which predicted that websites that have been designed for a particular target culture (e.g. Kuwait, or Egypt, or UK) using the cultural design claims will produce better usability results when tested by members of that particular target culture. Figure 3 shows a clear tendency for high usability by Kuwaiti participants in using their cultural design (design-A), but there is an exception to the hypothesis for Egypt and UK. Egyptian participants show high usability in using design-A, while UK participants have a usability score that is split between design-B and design-C. To further investigate the cause of this unexpected result, in the following section, the DA is used to identify which specific variables were affecting usability scores for each of the cultures.

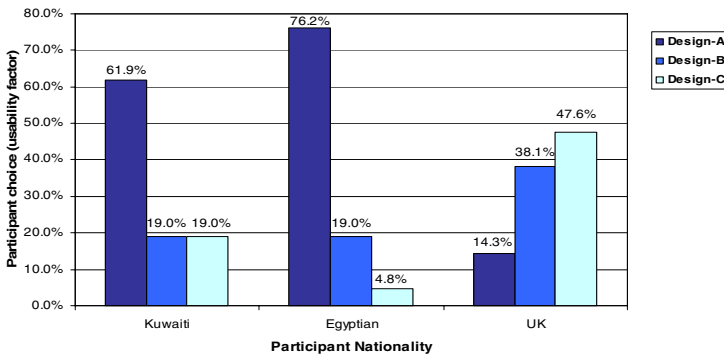


Fig. 3. The distribution graph for the usability scores according to culture and design

The Classification of the Three Designs Using DA Test. DA was performed with usability factor as the dependent variable, and the studied CMs (58 variables) as independent variables. This test provides two types of result. The first result is the classification of the three designs (A, B, and C) based on the usability factor for each

case study culture (to determine the usability level on different designs). The second result is in identifying the DMs, which cause usability improvements among specific national cultures as shown in Table 2. The DA results shows that the total validity of the proposed model is 100% for observations, which indicates that all cases were adequately categorized in all cultures. In addition, the visual graphs produced by the DA [9] show a divergence between the design type centroid points, which primarily discriminate between UK, Kuwaiti and Egyptian cultures. However, the design classification based on usability factor across cultures shows that design-A seems not to discriminate between Kuwaiti and Egyptian cultures. This confirms the results shown in Figure 3, which stresses that at the cultural usability level, Kuwaiti and Egyptian participants show some similarities in usable DMs. This indicates that, based on usability, Kuwait and Egypt could share design-A and that the UK site (design-C) should be redesigned to have cultural DMs from design-B, in addition to design-C DMs. Thus, study hypothesis (H2) is partially confirmed for Kuwaiti culture. However, to be sure of this conclusion we need to look at the DA results in more detail in order to determine which particular design factors were causing these usability effects. This will enable us to determine how to fine-tune the designs and modify the identified cultural design. The specific details of the DMs that affect these changes are identified and discussed in Table 2.

As can be seen from the summary DA results shown in Table 2, there is a clear tendency to identify specific DMs that are the main contributors to the observed participants' usability. Hence, H3 is confirmed for identifying the DMs for usability. This indicates the ability of the DA to identify the DMs that affect usability. These DMs are used as user-in-context based evidence in supporting or contradicting the cultural design claims. Reviewing the complete list of the usability DMs (see [9]) indicates that the shared DMs and cultures based on the cultural usability factor shows that there are more shared cultural usability DMs between Kuwait and Egypt, followed with Kuwait and UK. However, between Egypt and UK, there are no shared DMs. Again this confirms the relation between Kuwaiti and Egyptian cultures discussed earlier in sections 3.1 and 3.2.1. In addition, the DMs related to preferences and usability levels, the analysis shows that the identified DMs for preferences are higher than usability (see [9]). Furthermore, some usability markers appear to be different from preferences related DMs.

Table 2. Partial summary table for cultural usability DMs

CA	Claim	Design marker	KU	EG	UK
		Relationship Metaphors			
R6, R7	C16	National Metaphors (Design B)		H†	
		Navigation tools			
T4	C21	Drop-down Menu (complex navigation)	H†		H†
		Tree-view (complex navigation)			L†
		Drop-down field (complex navigation)	H†		H†
		Free-search (complex navigation)		H†	

Legend

† This symbol indicates that this DM affects usability for this particular culture (presenting a cultural-usability design). The result of this indicator is determined by performing DA.

4 Discussion and Conclusion

The Cross-Use data analysis was presented through two models. The first model is the cultural preferences model, which consists of the high level classification and DMs of cultural preferences (as shown in Section 3.1), and the second model is the cultural usability model, which consists of the high level classification and DMs of cultural usability (as shown in Section 3.2). Both models have different concepts that require various analysis techniques, which produce diverse results and significance levels. The cultural preferences model concept was to identify whether the participants' preferences for using the three designs are different, where the experiment shows there are significant differences. This proves that the experiment designs were able to classify cultures based on participants' preferences for the DMs, which at one level substantiates the experiment design and on the other level shows that there are cultural design differences. In addition, this model shows that a high number of the identified DMs are culturally preferred, which indicates that most of the DMs can be differentiated based on participants' preferences.

The next challenge here was to see whether the usage of culturally preferred DMs in local designs improves local design usability. This led to the development of the second model, which covers usability and was referred to as the cultural usability model. The cultural usability model was developed based on how the user performs the assigned six tasks (see Section 2.1), where the usability factor was developed to discriminate between the studied cultures. Based on this model, several issues were identified. The first issue shows that there is a high relation between culture and design usability using the three designs. This indicates that the three designs were able to identify a relation between culture and usability, which shows that at the classification level culture preferences are able to make usable designs. However, based on the most usable design related to culture, the results show that the Egyptian culture reflects design-A as the most usable design compared to the earlier expectation, which is design-B. In addition, the UK participants shared both design-C and design-B as they are the most usable designs (as shown in Figure 3). Therefore, the cultural DMs based on usability are not the same as the cultural design claims. These findings motivate the investigation of cultural usability DM.

Earlier, design preferences and usability were discussed to determine their differences. Then, during the experiment evaluation, these two issues were tested using a process to evaluate users. The question here is whether the websites that have been designed based on user cultural preferences are necessarily presenting usable design. The answer to this question helps in recognizing the sensitivity of the approach in collecting data that provides results to help in delivering usable design. The study of Evers and Day [3] uses the culturally extended Technology Acceptance Model (TAM), which uses the usability variables such as usefulness, ease of use, and satisfaction to determine the UI acceptance. They use questionnaires to collect users' preferences. Their study indicates that design preferences affect interface acceptance across cultures. In the Cross-Use experiment, the general view of the design classification based on the usability factor for each culture shows higher differences on cultural preferences than usability (see [9]). This proves that participants prefer design differently, but when they use the design, it shows more differences in usability than

originally expected. This highlights the complementary usage of the user-in-context evaluation in determining the usable cultural DMs.

Many website developers and evaluators use methods that assess user preferences aiming to create usable design. For example, the Cultural Markers [5], Website Audit [8], and user evaluation [10] using questionnaire based tools only are not sufficient in understanding and identifying the appropriate usability requirements. According to the results of Cross-Use experiment, as can be seen from Table 1, which presents user preferences CMs, and Table 2, which presents usability CMs, the comparison between the two markers indicates that the number of the identified markers in each type is different, and the identified markers based on preferences are not necessarily identified based on usability and vice-versa. The cultural usability model identifies fewer DMs than in the cultural preferences model. These prove that not all of the preferred DMs are necessarily usable DMs. Furthermore, the cultural usability DMs show that there are some DMs that are not shown to be preferred by the participants but are statistically proven to improve usability (e.g. Tree-view navigation DM in claim C21, as shown in Table 1 and Table 2). This suggests that research based on design preferences does not necessarily present the effects of usability as indicated by Constantine and Lockwood [4]. As a consequence, the results of such studies linking participants' preferences to design can be doubted, and this also affects the investigation of existing website design, as both adopt the same results. Therefore, the results obtained from users' preferences and usability should scale differently in supporting cultural design claims and in the later stages of the development of cultural design guidelines.

This conclusion strengthens the research results as they are obtained by evaluating both the cultural preferences and usability DMs. For the future research a detailed inspection method are expected to be used to analyse these results together with results of earlier research studies, which aims at developing evidence-based cultural design guidelines and recommendations.

References

1. MacKenzie, S.: Research Note: Within-subjects vs. Between-subjects Designs: Which to Use? Toronto, Ontario, Canada, <http://www.yorku.ca/mack/RN-Counterbalancing.html> (2002)
2. Pallant, J.: SPSS Survival manual. McGraw, New York, USA (2005)
3. Evers, V., Day, D.: The Role of Culture in Interface Acceptance. Human-Computer Interaction, Interact'97, London (1997)
4. Constantine, L., Lockwood, L.: Software for Use: A Practical Guide to the Models and Methods of Usage-Centred Design. Addison-Wesley, New York (1999)
5. Barber, W., Badre, A.: Culturability: The Merging of culture and usability. In: Proceedings of the 4th Conference of Human Factors and the Web (1998)
6. Marcus, A.: User Interface Design and Culture. In: Aykin, N. (ed.) Usability and Internationalization of Information Technology, pp. 51–78. Lawrence Erlbaum Associates, Inc, Mahwah, New Jersey (2005)
7. Bourges-Waldegg, P., Scrivener, S.A.R.: Applying and testing an approach to design for culturally diverse user groups. Interacting with computers 13, 111–126 (2000)

8. Smith, A., Dunckley, L., et al.: A process model for developing usable cross-cultural websites. *Interacting with computers* 16, 63–91 (2004)
9. Alostath, J.: *Culture-Centred Design: Integrating Culture into Human-Computer Interaction*. Doctoral Thesis, The University of York, UK (2006)
10. Evers, V.: *Cultural Aspects of User Interface Understanding: An Empirical Evaluation of an E-Learning website by International User Groups*. Doctoral Thesis, the Open University (2001)
11. Brace, N., Kemp, R., Rosemary, S.: *SPSS for psychologists: A guide to data analysis using SPSS for Windows*, vol. vii, p. 287. L. Erlbaum Associates, Mahwah, NJ (2003)
12. Rosson, M.B., Carroll, J.M.: *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*. USA (2002)
13. Alostath, J., Wright, P.: *Integrating Cultural Models into Human-Interaction Design*. Conference on Information Technology in Asia 2005 (CITA2005), Kuching, Sarawak, Malaysia (2005a)