CHAPTER 4

THE EXPERIMENT

4.1 Introduction

This chapter reports on the experiment conducted to determine whether or not subjective culture affects the usability of user interfaces. The chapter begins by identifying the objectives and hypotheses relating to the experiment in terms of the specific cultural dimensions selected for the study (section 4.2), followed by an overview of the experimental design (section 4.3). We then discuss the experiment methodology, in terms of the instruments used in the measurement of the key variables, the sample design and sampling method, the methods used for the collection, capture and analysis of the data, and the limitations of the experiment (section 4.4). The results of the experiment are presented next (section 4.5), followed by an analysis and interpretation of the results (section 4.6). The chapter concludes with the identification of the variables that could have influenced the results of the experiment.

4.2 Objectives and Hypotheses

Given the multitude of subjective cultural aspects and related guidelines identified in the literature, we felt it necessary to select one subjective cultural model in order to conduct a more focused study. Although the use of cultural models and Hofstede's [2001] cultural model in particular, has been criticized, we found substantial theoretical evidence (presented in Chapter 3) that this model influences usability and performance. As a result, the Hofstede model was used as the theoretical foundation for the experiment.

The objective of the experiment is therefore to establish empirical evidence that Hofstede's cultural dimensions affect the usability of computer-based systems. The following hypotheses, derived from Hofstede's cultural model, were identified and tested:

H1 Power distance will affect the usability of a computer-based system.

H2 Uncertainty avoidance will affect the usability of a computer-based system.

H3 Masculinity vs. femininity will affect the usability of a computer-based system.

H4 Individualism vs. collectivism will affect the usability of a computer-based system.

H5 Time orientation will affect the usability of a computer-based system.

4.3 Experimental design

As discussed in Chapter 3, a causal experimental design was used, in conjunction with primary data. The impact of Hofstede's [2001] cultural dimensions on usability was assessed by measuring the performance of users with different cultural profiles using different web site interfaces.

The independent variables were identified as the cultural profile of the test subjects and the cultural profile of the test interfaces. Given that usability is measured in terms of performance, the dependent variables were identified as speed, accuracy and satisfaction.

The experiment data were analyzed to determine whether a causal relationship existed between the users' performance, the users' cultural dimensions and the test interfaces' cultural fit. For example, it was expected that high power distant users using a high power distant site would achieve significantly different performance levels than when using a low power distant site.

4.4 Experiment Methodology

The primary research method used was a controlled experiment, in the form of a formal usability test, supported by the use of questionnaires. The questionnaires are discussed in detail in section 4.4.1.

Data on performance measures were collected quantitatively, using a test task instrument that comprised of test tasks and a satisfaction questionnaire. Interval measures were used to measure satisfaction, using a semantic differential, 5-point rating scale. Interval measures were also used to measure speed and accuracy. The methods used to collect data on performance measures are discussed in detail in section 4.4.2.

As the data obtained was quantitative in nature, statistical tests were used to analyse the data. Related samples t-tests were used to measure whether or not there were significant differences in accuracy, speed and satisfaction levels between users using an interface that corresponded to the subjects' cultural dimension side, compared to the same users using an interface with an opposing cultural dimension side. Independent samples t-tests were used to determine if there were significant differences in the test interfaces and test users. These data analysis methods are described in detail in section 4.4.3.

4.4.1 Measurement Issues

Three instruments were used in the experiment: one was used to determine the user profile of the test subjects, which included the subjective cultural profile of the test subjects as well as the other user characteristics specified in the user context of use (discussed in section 2.4.2.1). The second questionnaire was used to determine the cultural profile of the test interfaces, while the third questionnaire was used to measure the performance of the test subjects. These instruments are discussed next.

4.4.1.1 Subjective Cultural Profile of the Test Subjects

The test subjects used in the experiment needed to represent opposing sides of each cultural dimension, whilst being homogenous in terms of all other user characteristics. One instrument, namely the user profile questionnaire, was used to assess the cultural profile and homogeneity of the test subjects. This instrument is attached as Annexure B.

A Value Survey Model (VSM) was developed by Hofstede [2001] to assess the cultural profiles of the respondents in his studies. The VSM consists of 26 questions, 20 content and six demographic questions. The 20 content questions are used to calculate the five cultural dimensions, where four questions each are used for every dimension [IMIT, 2003]. As discussed in section 2.5.3.4, the VSM was validated through its use in a number of studies. Although the VSM was available for assessing the cultural profile of the test subjects, the test subjects were sourced from students with no working experience. Consequently the VSM had to be adapted to suit the test subjects' context. Questions were developed based on the questions used in the Value Survey Model and the cultural dimension characteristics described below [Marcus, 2000].

a. Power Distance

This dimension refers to the extent to which less powerful members of a society or group of people expect and accept unequal power distribution within that group [Marcus, 2000]. High power distant (HPD) people are afraid to express disagreement with people in authority such as bosses, parents and teachers. Low power distant (LPD) people have little difficulty in approaching and contradicting their superiors. Table 4.1 identifies the questions that were incorporated into the questionnaire as well as the answers that were expected from high and low power distant users.

Cultural Profile Question	HPD	LPD
If a lecturer says something that I disagree with, I will challenge the lecturer, during the lecture (26)	Disagree	Agree
If a lecturer says something that I disagree with, I will challenge the lecturer after the lecture (31)	Disagree	Agree
If a lecturer disagrees with the work that I have submitted, and I feel that I am in the right, I will take it up with the lecturer and stand up for my point of view (36)	Disagree	Agree
I prefer to discuss lecture material with tutors rather than with lecturers (39)	Agree	Disagree
I often discuss lecture material with my lecturers outside of lecture times (41)	Disagree	Agree

Table 4.1: Questions Used to Assess the Levels of Power Distance of the Test Subjects

b. Uncertainty Avoidance

Uncertainty avoidance is the way in which people cope with uncertainty and risk. High uncertainty avoidant (HUA) users tend to be emotional and aggressive, avoid ambiguous situations, prefer to work in a structured and predictable environment, and consider differences to be threatening and dangerous. It is also believed that high uncertainty avoidant users would prefer to work within a team environment, as this would serve as a support structure in times of uncertainty. In contrast, low uncertainty avoidant (LUA) users can accept that superiors do not have all the answers, that there may be more than one correct answer, and are curious about differences. Table 4.2 identifies the questions that were incorporated into the questionnaire as well as the answers that were expected from high and low uncertainty avoidant users.

Cultural Profile Question	HUA	LUA
I am more comfortable in a learning environment with structured timetable slots and precise learning objectives than in an open-ended learning environment (28)	Disagree	Agree
I have no problem in proceeding with a task even if the objectives are initially not clearly defined (for example, proceeding with an assignment but I don't know how it will be marked) (33)	Disagree	Agree
I would prefer to work on the Major Project on my own, rather than as a group, if there would be the same amount of work for me if I worked on the project as a group or on my own (37)	Disagree	Agree
Unfamiliar situations make me feel uncomfortable (38)	Agree	Disagree
I think that the correct answer is more important than an original / creative answer (44)	Agree	Disagree

Table 4.2: Questions Used to Assess the Levels of Uncertainty Avoidance of the Test Subjects

c. Masculinity vs. Femininity

This cultural dimension refers to gender roles, not physical characteristics, and is primarily characterized by the levels of assertiveness or tenderness in the user. Masculine (MAS) users tend to be assertive, competitive and tough. Their work goals include high earnings,

recognition, advancement and challenge. Feminine (FEM) users focus on home, children and people. Their work goals include good relations with supervisors, peers and subordinates, good living and working conditions with sense of security. Table 4.3 identifies the questions that were incorporated into the questionnaire as well as the answers that were expected from masculine and feminine users.

Cultural Profile Question	MAS	FEM
Competing with my fellow students academically or otherwise is not important to me (29)	Disagree	Agree
A lecturer who is friendly and sociable is a better lecturer than one who has a strong academic reputation (34)	Disagree	Agree
Salary and promotions are more important to me than caring and social roles in my job (42)	Agree	Disagree
It is more important for me to get the recognition that I deserve for the work that I do, rather than to work with people who cooperate well with one another (43)	Agree	Disagree
It is more important to me to have a challenging job than a job that provides me with good working conditions (45)	Agree	Disagree

Table 4.3: Questions Used to Assess the Levels of Masculinity and Femininity of the Test Subjects

d. Individualism vs. Collectivism

This dimension relates to the role of the individual and the group, and is characterized by the level of ties between an individual in a society [Hoft, 1996]. Individualist (IND) users are expected to look after themselves and their immediate family, but no one else. They value personal time, freedom and challenge, material rewards, honesty and truth, talking things out, maintaining self-respect, and the right to privacy and personal opinion. In contrast, collectivist (COL) users are integrated into strong, cohesive groups that protect them in exchange for unquestioning loyalty. Collectivists value training and skills, and group achievement rather than personal recognition. Harmony is valued more than truth and honesty. They are comfortable with an invasion of privacy and restrictions on personal opinions. Table 4.4 identifies the questions that were incorporated into the questionnaire as well as the answers that were expected from high individualist and collectivist users.

Cultural Profile Question	COL	IND
Social acceptance is more important to me than self-respect (27)	Agree	Disagree
When doing a project as a group (eg the Major Project), each member should get the same mark for that project, rather than each member getting assessed individually (32)	Agree	Disagree
I would prefer to work on the Major Project on my own, rather than as a group, if there would be the same amount of work for me if I worked on the project as a group or on my own (37)	Disagree	Agree

Table 4.4: Questions Used to Assess the Levels of Individualism and Collectivism of the Test Subjects

e. Time Orientation

This cultural dimension relates to people's concern with the past, present and future. In essence, short-term oriented (STO) people are concerned with the past and the present, while long-term oriented (LTO) people are concerned more with the future [Hoft, 1996]. Long-term oriented users believe that a stable society requires unequal relations, and that older people and men have more authority than younger people and women. They value trying to acquire skills and education, working hard and being frugal. They are prepared to persevere and display a lot of patience in understanding new things. In contrast, short-term oriented users believe in an equality of relationships, and emphasize individualism. They value reciprocity of favours, gifts and greetings, and the ability to achieve quick results. Table 4.5 identifies the questions that were incorporated into the questionnaire as well as the answers that were expected from short-term and long-term oriented users.

Cultural Profile Question	LTO	STO
If I do a favour for someone, I expect that person to return the favour when I need it (eg If I give a lift to a friend, I expect that friend to give me a lift when I need one) (30)	Disagree	Agree
I believe in living my life for the moment, rather than planning for the future (35)	Disagree	Agree
When I am learning something new and difficult, such as a new computer program, I persevere until I understand it (40)	Agree	Disagree

Table 4.5: Questions used to Assess the Levels of Time Orientation of the Test Subjects

4.4.1.2 User Characteristics

Questions relating to objective culture and user characteristics were incorporated into the user profile questionnaire to identify any variables other than the cultural dimensions that may have affected the outcome of the test [Olivier, 2004], These questions were based on prior research done by Mayhew [1992], Shneiderman [1998], and Rubin [1994], and incorporated those characteristics that were deemed relevant to the experiment: These are illustrated in the Table 4.6 (the corresponding question number in the questionnaire is indicated in brackets after the criterion).

User Characteristics	Criterion
Objective Culture	Home Language (3)
	Racial Group (5)
Knowledge and Experience	Degree registered for (1)
	General level of computer experience (8)
	Level of typing skill (9)
	Prior experience on test interfaces (11)
	Length of time using a computer (12)
Job and Tasks	Frequency of Computer Usage (13)

Table 4.6: Questions Relating to User Characteristics and Objective Culture

4.4.1.3 Test Interfaces

The test interfaces were also required to display characteristics appropriate to the cultural dimensions in Hofstede's model. A Website Evaluation Checklist was developed for this purpose, based on a set of guidelines proposed by Marcus [2002] for accommodating Hofstede's cultural dimensions into the design of user interfaces. The checklist is attached as Annexure C, and the design guidelines for each cultural dimension are discussed below.

a. Power Distance

Interfaces that display high power distance characteristics should provide highly structured access to information, prominence should be given to leaders, security measures should be both explicit and enforced and there should be a strong focus on authority. The opposite holds true for low power distant sites. The recommended implementation of these characteristics per interface component is presented in Table 4.7.

Component	High Power Distance	Low Power Distance
Metaphors	Institutions, buildings and objects with clear hierarchy such as schools, government buildings and monuments	Institutions, buildings and objects with equality, such as play/games and public spaces
Conceptual Models	Reference data with no relevance ranking	Less structured data with relevance
Navigation Methods	Restricted access and choices, authentication, passwords, prescribed routes	Open access, multiple options, sharable paths
Interaction Devices	Severe error messages, wizards or guides lead usage	Supportive error messages, cue cards
Appearance	Images of leaders and nations, official music such as anthems, formal speech	Images of people and daily activities, popular music, informal speech

Table 4.7: Interface Component Characteristics for Dower Distance

b. Uncertainty Avoidance

Interfaces that display high uncertainty avoidance characteristics should focus on the prevention of user error by providing minimal menu options, simple and descriptive help facilities, and a navigation structure that is focused on preventing users from getting lost. Colours, sounds and images should be used to reinforce the messages. In contrast, low uncertainty avoidant interfaces should encourage user exploration; provide many menu options, and use colours, sounds and images to provide additional information. The recommended implementation of these characteristics per interface component is presented in Table 4.8.

Component	High Uncertainty Avoidance	Low Uncertainty Avoidance
Metaphors	Familiar, clear references to daily life, representation	Novel, unusual references, abstraction
Conceptual Models	Simple, clear articulation, limited choices, binary logic	Tolerance for ambiguousness, complexity, fuzzy logic
Navigation Methods	Limited options, simple and limited controls	Multiple options, varying and complex controls
Interaction Devices	Precise, complete, detailed input and feedback of status	General, limited or ambiguous input and feedback of status
Appearance	Simple, clear consistent imagery, terminology and sounds, highly redundant coding	Varied, ambiguous, less consistent imagery, terminology and sounds

Table 4.8: Interface Component Characteristics for Uncertainty Avoidance

c. Masculinity vs. Femininity

Interfaces that are oriented towards the masculine side of this dimension should be focused on allowing for quick results for limited tasks. The navigation structure should support user exploration and control. The content should be suggestive of a challenge for the user to master something, and cater for explicit distinctions between genders and age groups. Graphics and animations should be used for utilitarian purposes. In contrast, feminine oriented interfaces should use aesthetic appeal and poetry as a way of gaining users' attention. There is a blurring of gender roles. In particular, feminine oriented interfaces should support mutual cooperation and the exchange of ideas and support. The recommended implementation of these characteristics per interface component is presented in Table 4.9.

Component	Masculinity	Femininity
Metaphors	Sports oriented, competition oriented, work oriented	Shopping carts, family oriented
Conceptual Models	Work/business structures, high level executive views, goal oriented	Social structures, detailed views, relationship oriented
Navigation Methods	Limited choices, synchronic	Multiple choices, multi-tasking, polychronic
Interaction Devices	Game oriented, mastery oriented, individual oriented	Practical, function oriented, cooperation oriented, team oriented
Appearance	Masculine colors, shapes, sounds	Feminine colors, shapes, sounds, acceptance of cuteness

Table 4.9: Interface Component Characteristics for Masculinity vs. Femininity

d. Individualism vs. Collectivism

Individualist interfaces should use images of materialism and consumerism to denote success, and youth, action and individuals to gain the users' attention. The content should be focused on personal achievement, new and unique products and concepts, and contain or encourage controversy and personal opinions. Users should not be required to provide personal information. In contrast, collectivist sites should use images of the achievement of socio-political agendas to denote success, and experienced, aged leaders and groups of people to gain the user's attention. The content should be focused on group achievement, history and tradition, and contain official slogans while discouraging personal opinions. The recommended implementation of these characteristics per interface component is presented in Table 4.10.

Component	Individualism	Collectivism
Metaphors	Action oriented tools	Relationship oriented
Conceptual Models	Product or task oriented	Role oriented
Navigation Methods	Individual paths, popular choices, celebrity choices, stable across roles, customizable	Group oriented, official choices, changes per role
Interaction Devices	Keyword searches, active oriented, multiple devices, customizable	Limited, office devices, role driven
Appearance	Images of products, people, low context, hyperbolic, dynamic speech, market driven topics, imagery and language, customizable, direct and active verbs	Images of groups, organizations; images of roles; high context; official, static terminology,; institution driven topics, imagery, language; passive verbs

Table 4.10: Interface Component Characteristics for Individualism vs. Collectivism

e. Time Orientation

Short-term oriented user interfaces should be structured in a way that allows users to complete tasks quickly. Rules should be used to verify the credibility of information, and information content should be based on truth and certainty of beliefs. In contrast, long-term oriented

interface navigation style and content can be more complex, as users will persevere until they gain an understanding. Long-term oriented websites should contain content that is of practical value, and can use relationships to verify the credibility of the information. The recommended implementation of these characteristics per interface component is presented in Table 4.11.

Component	Short Term Orientation	Long Term Orientation
Metaphors	Interchangeable roles, jobs, objects	Stable family, paternalistic: Father, Mafia, Chinese state businesses, IBM in 1950s
Conceptual Models	Liberty: social incoherence, social irresponsibility, efficiency	Interchangeable roles, jobs, objects
Navigation Methods	Bread-crumb trails, taxonomies; quick-results; action-oriented	Tolerance for long paths, ambiguity; contemplation-oriented
Interaction Devices	Distance communication accepted as more efficient; anonymous messages tolerated; conflict tolerated, even encouraged; performance-critical communication	Preference for face-to-face communication, harmony; personalized messages; more links to people; live chats; interaction as 'asking'
Appearance	Minimal and focused images; short borders, lines, edges; concentration on showing task or product	Cultural markers: flags, colors, national images; soft focus; warm, fuzzy images; pictures of groups inviting participation, suggestions of intimacy and close social distance

Table 4.11: Interface Component Characteristics for Time Orientation

4.4.1.4 Performance

In keeping with the standard usability metrics discussed in Chapter 2, performance was measured in terms of accuracy, speed and satisfaction. Accuracy and speed were measured by providing test subjects with a set of test tasks, that contained questions relating to different websites. The answers to these questions were to be found in the content of each web site. Satisfaction was measured by test subjects completing a questionnaire indicating their level of satisfaction after using each web site.

The test tasks and satisfaction questionnaire were combined into a Test Task Instrument that was handed out to each test subject at the start of each experiment session, and is appended as Annexures C and D. Following the method used by Spool et al. [1999], the test tasks were designed so that each answer comprised a single fact, and there was only one correct answer. The questions used to measure satisfaction were also developed based on the satisfaction questionnaires used by Spool, et al. [1999].

4.4.2 Sample Design and Sampling Method

The objectives of the experiment necessitated that samples of test subjects and test interfaces be identified. The sample subjects were required to display characteristics pertaining to one

side of each cultural dimension being tested. The test interfaces were also required to display these characteristics. Consequently, we discuss the sample design and sampling techniques used to identify test subjects and test interfaces below.

4.4.2.1 Test Subjects

The test subjects were selected from the group of 120 students enrolled for the third level course in Business Information Systems (BIS) at the University of KwaZulu-Natal, Durban. This group was selected with the expectation that they could share a similar user profile in terms of the same educational background and similar computer skills. This expectation was based on the premise that the students have all been exposed to the same first and second level courses, as offered by the University of KwaZulu-Natal as pre-requisite courses for the third level course. These courses include a first and second level course in (BIS). In addition, care was taken to ensure that the test subject group was representative of gender, and that the ages of the subjects were similar.

All candidates were required to complete the User Profile Questionnaire. The questionnaire was handed out during a lecture period and time was set aside for the students to complete the questionnaire during the lecture period. On completion, the questionnaires were collected during the lecture period.

To control for variables other than cultural dimensions, test subjects were selected by filtering the candidate population on the following basis (the numbers in brackets refer to the question numbers in the cultural profile questionnaire):

(a) Degree registered for (1) – only those students registered for the BCom / BBus Sci degree in Business Information Systems, or the BSc degree in Computer Science.
(b) Home Language (3) – only those students whose home language is English.
(c) Racial Group (5) – only Asian students were selected, as this racial group had the largest number of students that were homogenous in terms of the other filtering criteria.
(d) General level of computer experience (8) – moderately high or high.
(e) Level of typing skill (9) – moderate or high.
(f) Prior experience on test interfaces (11) – test subjects that had previously purchased anything from Barnes and Noble were filtered out.
(g) Length of time using a computer (12) – more than four years.
(h) Frequency of computer usage (13) – more than once a week.
(i) Colour blindness (21) – users that were colour blind were filtered out.
(j) Other visual impairments (22) – users that had visual impairments were filtered out.
(k) Physical handicaps (23) – users that had physical impairments that could have affected their speed of movement were filtered out.

The cultural dimensions of the selected test subjects are summarized in Table 4.12.

Cultural Dimension	Side	No	%	Total
Power Distance	HPD	18	32	57
	LPD	39	68	
Uncertainty Avoidance	HUA	30	56	54
	LUA	24	44	
Masculinity / Femininity	MAS	18	32	57
	FEM	39	68	
Individualism / Collectivism	IND	37	69	54
	COL	17	31	
Time Orientation	LTO	52	93	56
	STO	4	7	

Table 4.12: Number of Users Selected for Each Side of Each Dimension

The majority of users scored in the low levels of each dimension. This is depicted graphically in the form of histograms attached as Annexure E.

It was interesting to note that 93% of the test users were long-term oriented. The majority of users were also low power distant, feminine, high uncertainty avoidant and individualists. In contrast, Hofstede's study [1991] found that the dominant cultural profile for South Africa was low power distant, individualist, masculine and low uncertainty avoidant (no score was obtained for time orientation). This comparison is summarized in Table 4.13.

Dominant Cultural Profile of	Power Distance	Uncertainty Avoidance	Masculine / Feminine	Individual / Collective	Time Orientation
Test Subjects	Low	High	Feminine	Individualist	Long-term
South Africa	Low	Low	Masculine	Individualist	N/A

Table 4.13: Comparison of Dominant Cultural Profiles

The dominant cultural profile of the test subjects was therefore different to the dominant cultural profile of South Africa identified in Hofstede's survey. One possible explanation for the dominance of high uncertainty avoidance in the test subjects could be attributable to the fact that the test subjects were all students. One possible reason for the difference in dominance for the masculine / feminine dimension could be that Hofstede's study indicated a low level of masculinity for South Africa, suggesting that there is no dominant side of this dimension for this country. However, it is equally possible that the survey did not include a sufficient representation of the diverse cultures in South Africa. The survey was conducted during the late 1960s and early 1970s, during which time management levels in South Africa were dominated by white males.

4.4.2.2 Test Interfaces

The websites that were used as test interfaces for the experiment are listed in Table 4.14.

Dimension	Side	Site	Side	Site
Power Distance	High	University of Malaysia (www.uum.edu.my)	Low	University of South Africa (http://osprey.unisa.ac.za)
Uncertainty Avoidance	High	Likouris Travel (www.travelagent.gr)	Low	Singapore Airlines (www.singaporeair.com.sg)
Masculinity vs. Femininity	MAS	Sony (Sweden) (www.sony.se)	FEM	IBM (USA) (www.ibm.com)
Individualism vs. Collectivism	IND	US National Park (www.nps.gov/glba)	COL	Costa Rican National Park (www.tourism.costarica.com)
Time Orientation	Long-term	Singapore Tourism Cybrary (www.cybrary.com.sg)	Short-term	Barnes and Noble (USA) (www.bn.com)

Table 4.14: Test Interfaces Identified from the Initial Evaluation

Five sets of websites that display each of the required cultural dimensions were identified from the existing literature [Marcus, 2001]. These are listed in Table 4.15.

Dimension	Side	Web Site	Side	Web Site
Power Distance	HPD	Malaysian University	LPD	Dutch Educational
Individualism vs. Collectivism	IND	US National Park	COL	Costa Rican National Park
Masculinity vs. Femininity	MAS	Excite.com (Japan)	FEM	Excite.com (Sweden)
Uncertainty Avoidance	HUA	Sabena Airlines (Belgium)	LUA	British Airways (UK)
Time Orientation	STO	Siemens (Germany)	LTO	Siemens (China)

Table 4.15: Websites Reflecting Cultural Dimension Characteristics [Marcus, 2000]

It was originally intended to make use of these websites for the experiment, however, five of the websites were not appropriate for the test tasks, as the content was not English, and one of the websites had been discontinued as a result of the company declaring insolvency. Furthermore, one of the websites became unavailable while conducting the experiment, and as the subjects had already seen the questions, could not be used during a later session. Table 4.16 lists the cultural dimensions that required replacement websites.

Dimension	Side	Web Site	Reason
Power Distance	LPD	Dutch Educational	Content not in English
Masculinity vs. Femininity	FEM	Excite.com (Sweden)	Content not in English
Masculinity vs. Femininity	MAS	Excite.com (Japan)	Content not in English
Uncertainty Avoidance	HUA	Sabena Airlines (Belgium)	Discontinued
Uncertainty Avoidance	LUA	British Airways (UK)	Unavailable during experiment
Time orientation	STO	Siemens (Germany)	Content not in English
Time orientation	LTO	Siemens (China)	Content not in English

Table 4.16: Reasons for Unsuitability of Websites Identified by Marcus [2000]

As a result, only three of the test interfaces that were identified for use are specified and evaluated by Marcus [2000]. These are the University of Malaysia, the US National Park and the Costa Rican National Park websites. Although the Malaysian and Costa Rican sites were redesigned since Marcus's evaluation in 2000, the redesigns took place after the experiment was conducted.

Replacement websites needed to be identified for seven of the test interfaces. These were identified through the following process:

(1) **Initial evaluation:** Countries with high and low indices for the various cultural dimensions were identified from Hofstede's [2001] survey. Potential websites were then identified from those specific countries and evaluated in terms of the characteristics relevant to each cultural dimension

(2) **Comparative Evaluations:** Fourth level students (registered for the Hons BCom (IST) degree) were given the seven websites to evaluate independently as part of their coursework. The evaluations done by the researcher and the fourth level students were compared. This comparison was successful for all seven websites.

(3) **Retro-Evaluation:** After the experiment had been conducted, a new literature source became available that proposed a set of guidelines for accommodating Hofstede's cultural dimensions into the design of user interfaces [Marcus, 2002]. A website evaluation checklist was developed based on this new literature source, which was used to retro-evaluate the cultural fit of the test interfaces. The retrospective evaluation of the web site evaluated the interfaces' cultural profile in terms of the five components that comprise an interface [Marcus, 2002], namely metaphors, mental model, navigation, interaction and appearance. The retrospective evaluation resulted in surprising results for the Sony, IBM and Singapore tourism sites, when compared to their respective countries' cultural dimension indices in Hofstede's analysis [Marcus, 2000].

- Sweden is the highest scoring country for femininity, yet the Sony (Sweden) website displays many more masculine characteristics than feminine ones (see Annexure A-3.1).
- The USA scored as a medium masculinity country, yet the Barnes and Noble (USA) website displays a marked set of feminine characteristics (see Annexure A-3.2).
- Singapore had a score of 48 in the time orientation index, indicating that it was a short-term oriented country. However, the evaluation of the Tourism website found it to be long-term oriented (ref Annexure A-5.1).

These findings could be attributed to the suggestion that cultural boundaries do not necessarily coincide with national boundaries [Duncker, 2002]. National states often comprise multiple cultures and ethnicities, of which South Africa's 'Rainbow Nation' is a prime example. Therefore, it is highly likely that the websites were developed by designers with a cultural profile different to that of each of the countries.

These evaluations were then submitted to Marcus for confirmation [Marcus and Baumgartner, 2002, Marcus, 2002b]. Marcus and Baumgartner [2002] agreed with the overall findings of the retrospective evaluations of all the sites except for the two used for the masculinity vs. femininity dimension. They point out that in comparison to the IBM (Sweden) site, the IBM (USA) site is not strongly feminine in its characteristics. Also, the Sony (Sweden) site is more feminine than its USA counterpart. This argument is accepted. However, in comparing the IBM (USA) site to the Sony (Sweden) site, the IBM (USA) site does exhibit more feminine characteristics than the Sony (Sweden) site. Consequently, for the purposes of this research, user performance on the IBM (USA) site will be analysed in terms of the feminine side of the cultural dimension, and the Sony (Sweden) site performance will be analysed in terms of the masculine side.

Marcus [2002b] suggested that it is possible that only some of the screens, or some of the components of the websites, displayed the properties relevant to the identified cultural dimension side. He also pointed out that the chosen sites had varying degrees of design expertise, which could lead to misleading results. These suggestions confirmed the need to perform the independent samples t-tests for site and user differences and the paired samples t-tests for usability equality.

4.4.3 Data Collection Methods

Data were collected in terms of the performance achieved by the test subjects using the test interfaces. This was done in five sessions, one cultural dimension being tested per session. The reason for this was that the total amount of time required to complete the tasks for the ten websites was one-and-a-half hours. It was believed that the test subjects would become tired, impatient and bored, possibly leading to incorrect results.

Measurements were recorded of each user's speed, accuracy and satisfaction when using an interface that corresponded to the user's cultural dimension side. The same measurements were recorded of the user when using an interface with the opposing cultural dimension side; therefore a within-subjects design was used in that the same test subjects were required to access both websites. Table 4.17 illustrates the measures recorded for accuracy, speed and satisfaction, for each of the cultural dimensions during each of the five sessions.

User Profiles	**Website Profiles**									
	Session 1		Session 2		Session 3		Session 4		Session 5	
	HPD	LPD	HUA	LUA	MAS	FEM	IND	COL	LTO	STO
HPD	X	X								
LPD	X	X								
HUA			X	X						
LUA			X	X						
MAS					X	X				
FEM					X	X				
IND							X	X		
COL							X	X		
LTO									X	X
STO									X	X

Table 4.17: Measures Recorded for Accuracy, Speed and Satisfaction

Interval measures were used to measure accuracy and speed. The score for accuracy was calculated as illustrated in Table 4.18.

No of questions answered	No of answers that were correct	Score	Score as a Percentage
2	2	2	100
2	1	1	50
2	0	0	0
1	1	2	100
1	0	0	0
0	0	0	0

Table 4.18: Accuracy Scoring Method

For example, if the user answered both questions, and both answers given were correct, his final score for accuracy, expressed as a percentage, was 100%. If the user answered only one question, and the answer given was correct, his score for accuracy was 100%. This calculation was used to avoid accuracy measures being influenced by speed.

Speed was recorded as amount of time that each user spent using each website, regardless of the number of answers found, or the number of accurate answers given.

The questions in the satisfaction questionnaire are closed-ended, use interval measures and a semantic differential, 5-point Likert scale. This method has been used by other researchers such as Shneiderman [1998], Spool et al. [1999] and Hofstede [2001]. The scoring method used was dependent on whether the question was positively or negatively phrased, and is reflected in Table 4.19.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Positively phrased	2	1	0	-1	-2
Negatively phrased	-2	-1	0	1	2

Table 4.19: Scoring Method for Satisfaction

Subjects were required to access each web site, find and record the answer to both questions on the test task instrument, and then complete the satisfaction questionnaire. Subjects were given a time limit within which to complete the tasks for each web site, based on a pilot study that was done prior to the experiment. Overhead transparencies displaying the time at one-minute intervals were used to facilitate this process. The test tasks were performed in the University of KwaZulu-Natal's computer laboratories. Fourth level BCom (IST) students were used to assist in moderating the experiment.

4.4.4 Data Analysis

The use of controlled experiments using quantitative data requires that the hypotheses stated in section 4.3 be converted into a format that was conducive to statistical testing. This is best described by way of example. The power distance dimension is used for this purpose. The primary hypothesis ($H1_1$) states that 'power distance will affect the usability of a computer-based system'. This is a non-directional hypothesis as it can be retained whether the cultural dimension increases or decreases usability. The null hypothesis ($H1_0$) is therefore that power distance does not affect usability.

The primary hypothesis ($H1_1$) can be accepted if the accuracy, speed or satisfaction levels of high power distant users, when using a high power distant interface, are significantly different to the levels attained when using a low power distant interface. Similarly, the hypothesis can be accepted if the accuracy, speed or satisfaction levels of low power distant users, when using a low power distant interface, are significantly different to the levels attained when using a high power distant interface. The null hypothesis ($H1_0$) should be retained if the levels obtained when using the two different interfaces, by the same users, are the same. The same logic holds

true for the primary hypotheses about the other cultural dimensions.

Consequently, the scores achieved for accuracy, speed and satisfaction on each of the two websites (M1 and M2) in each set of test interfaces were compared against each other to determine if there was a significant difference between them. Table 4.20 illustrates the comparisons done for each set of measurements for each cultural dimension.

Cultural Dimension	**Accuracy**		**Speed**		**Satisfaction**	
	M1	M2	M1	M2	M1	M2
	User/Interface	User/Interface	User/Interface	User/Interface	User/Interface	User/Interface
Power Distance	HPD/HPD	HPD/LPD	HPD/HPD	HPD/LPD	HPD/HPD	HPD/LPD
	LPD/HPD	LPD/LPD	LPD/HPD	LPD/LPD	LPD/HPD	LPD/LPD
Uncertainty Avoidance	HUA/HUA	HUA/LUA	HUA/HUA	HUA/LUA	HUA/HUA	HUA/LUA
	LUA/HUA	LUA/LUA	LUA/HUA	LUA/LUA	LUA/HUA	LUA/LUA
Masculinity vs. Femininity	MAS/MAS	MAS/FEM	MAS/MAS	MAS/FEM	MAS/MAS	MAS/FEM
	FEM/MAS	FEM/FEM	FEM/MAS	FEM/FEM	FEM/MAS	FEM/FEM
Individualism vs. Collectivism	IND/IND	IND/COL	IND/IND	IND/COL	IND/IND	IND/COL
	COL/IND	COL/COL	COL/IND	COL/COL	COL/IND	COL/COL
Time Orientation	LTO/LTO	LTO/STO	LTO/LTO	LTO/STO	LTO/LTO	LTO/STO
	STO/LTO	STO/STO	STO/LTO	STO/STO	STO/LTO	STO/STO

Table 4.20: Measure Comparisons

Each hypothesis was tested separately using a different experiment session. During each session, therefore, only one independent variable (the cultural dimension), containing two treatments (each side of the cultural dimension), was used. A within-subjects design was used in that the same test subjects were required to access both websites. To test each hypothesis, it was necessary to assess the differences in scores attained by users using an interface displaying a matching cultural dimension side, compared to the same users using an interface displaying the opposing cultural dimension. Therefore, the group mean of performance of, for example collectivist users using a collectivist site was compared to the group mean of the same users using an individualist site. Consequently, for each hypothesis, two group means were compared, requiring the use of bivariate statistics in the form of either the t-test or ANOVA. As only two treatments were used, the t-test was chosen. Finally, as the two groups comprised of the same test subjects, a related samples t-test was used to measure the differences in performance.

By analyzing and interpreting the results of the experiment data on the basis of the above hypotheses, it would have been possible to determine not only whether each side of each

cultural dimension affects performance, but also whether the dimension side affects any one of the three measures of performance of accuracy, speed or satisfaction, or some combination thereof. If the cultural dimension significantly affected any of the performance categories, then it would have been concluded that the cultural dimension affects user performance. In addition, it was intended to determine which cultural dimensions had the greatest impact on performance, so that the dimensions could be ranked in order of importance for cost effective user interface design. The point-biserial coefficient [Heiman, 1996] was used to determine the cultural dimension (or side) that had the most impact on the three performance categories. This coefficient was only calculated for each significant difference found in the data.

It was noted during the analysis that variables other than cultural dimensions in both the sites and the user groups could have affected performance levels. For example, one of the two sites in a set could have been superior in some way, or one of the two sets of users could have, by chance, been performing better generally. These additional variables were controlled for by accepting the above primary hypotheses only if, in addition to significant differences being found as explained above, *no* differences were found between the sites or between the user groups.

As a result, four different t-tests were performed for each cultural dimension's set of measures:

(1) **Paired samples t-test:** This test was used to determine whether a user, using an interface with the corresponding side of the cultural dimension, performs better than when using an interface with the opposing side of the dimension. The two sets of measurements for each cultural dimension side (as illustrated in Table 4.20) were compared to ascertain whether the user performed significantly differently when using the corresponding interface than when using the opposing interface. If a significant difference was found ($\mu_D <> 0$), then independent samples t-Tests were performed on the two sites and the two user groups.

For example, the measure of accuracy for High power distant users using a High power distant site was compared to the measure of accuracy for High power distant users using a Low power distant site.

(2) **Independent samples t-test (Site):** This test was used to determine whether one of the sites was generally a 'better' site than the other. This was done by determining the average score achieved by all users using the first site and comparing it to the average score achieved by all users on the second site. If a significant difference was found ($\mu_S <> 0$), then it was concluded that one of the sites was better than the other, and therefore the increase in performance could be attributed to variables on the sites other than that cultural dimension.

(3) **Independent samples t-test (User):** This test was used to determine whether one set of users was generally a 'better' set of users than the other. This was done by determining the

average score of users of one side of the dimension using both sites, and comparing it to the average score of users of the other side of the dimension. Again, if a significant difference was found (μ_U <> 0), then it was concluded that there were variables in the test subject groups, other than culture, that were causing increased performance.

(4) **Paired samples t-test (Usability):** To confirm the findings of the independent samples t-Tests done on the users and the sites, an additional paired samples t-test was performed on the data. The data was arranged to compare the difference in scores between

- all users using an interface with the same dimension, and
- all users using an interface with the opposing dimension.

Where the sample size of users with one side of the dimension was higher than the sample size of users with the opposing dimension, a random sample of the higher number was taken, equivalent to the smaller number in the opposing side.

Because the same number of users was using the potentially superior site and the potentially inferior site, the usability difference between the sites should be nullified, Therefore, if a significant difference was *not* found, then the test was seen to support the findings of the independent sample t-tests described above.

Based on the above four tests, the hypotheses stated in section 4.2 above can be decomposed into the following set of secondary hypotheses for each performance measure for each side of each cultural dimension:

$$H_0 : \mu_D = 0 \text{ AND } \mu_S <> 0 \text{ AND } \mu_U <> 0$$

$$H_1 : \mu_D <> 0 \text{ AND } \mu_S = 0 \text{ AND } \mu_U = 0$$

(where

μ_D denotes the variance of the scores between the measures M1 – M2,

μ_S denotes the variance of the mean scores between either of the sites in a set of sites, and

μ_U denotes the variance of the mean scores between the test subject groups).

4.4.5 Shortcomings and Sources of Error

The primary limitation of the experiment relates to the validity of the design guidelines used to identify test interfaces. The validity of cultural user interface design guidelines are questionable because they have not been empirically tested with users from various countries, and because no other factor of web design has been taken into consideration [Jagne et al., 2004]. Additional sources of error potentially stem from the user profile: although the questions were tested

through a pilot study and found to be reliable, it is still possible that errors exist, which could result in an inaccurate assessment of the cultural profile of the users. In addition, only those user characteristics believed to have the most influence on usability were controlled for. It is therefore possible that the other user characteristics may cause unexpected variances in performance.

4.5 Results

The performance of more than 50 test subjects for four of the five cultural dimensions were measured and compared. Due to the very small sample size of short-term oriented users found, it was not possible to analyze the results for the time orientation dimension. The results of the four statistical tests performed on the data obtained on the measures for each cultural dimension are presented below.

4.5.1 Impact of Power Distance on Usability

There were no significant differences in the accuracy or satisfaction levels achieved. Significant differences in speed occurred within the low power distant user group, but not within the high power distant user group. This is illustrated in Table 4.21.

Dimension Side	Accuracy	Speed	Satisfaction
HPD/HPD vs. HPD/LPD T crit = +-2.1098	t stat =0.223	t stat = -1.975	t stat = 0.327
LPD/LPD vs. LPD/HPD T crit = +- 2.024	t stat = 0.703	t stat = 4.376 $r_{pb}^2 = 0.34$	t stat = -1.619
Differences in site (HPD site vs. LPD site)	N/A	t stat = -4.526 t crit = +- 1.982 $r_{pb}^2 = 0.157$	N/A
Differences in user groups (HPD users vs. LPD users)	N/A	t stat = 0.06 t crit = +-1.995	N/A
Control for Usability (same dimension side for user and website vs. different dimension side)	N/A	t stat = 0.249 t crit = +- 2.030	N/A

Table 4.21: Summary of Findings for Power Distance

It was noted that the difference within the low power distant user group was positive, indicating that low power distant users using the low power distant site took longer to complete the tasks than when using the high power distant site. The independent samples t-test indicated that irrespective of the user's side of the cultural dimension, it took longer to complete the tasks overall using the low power distant site than when using the high power distant site. This was

confirmed by the lack of significant results found in the paired samples t-test used to control for usability. No significant difference was found between the two user groups.

4.5.2 Impact of Uncertainty Avoidance on Usability

The only *insignificant* difference found at the 95% level in the comparisons was in the accuracy scores between low uncertainty avoidant users using a low uncertainty avoidant site compared to the same users using a high uncertainty avoidant site. This is reflected in Table 4.22. However, this difference fell just short of being significant in terms of the t-crit value, and could be accepted at the 94% level.

Dimension Side	Accuracy	Speed	Satisfaction
HUA/HUA vs. HUA/LUA	t stat = 2.408	t stat = -13.581	t stat = 9.518
t crit = +- 2.045	r_{pb}^2 = 0.167	r_{pb}^2 = 0.864	r_{pb}^2 = 0.758
LUA/LUA vs. LUA/HUA	t stat = - 2.000	t stat = 9.239	t stat = -10.585
t crit = +- 2.069		r_{pb}^2 = 0.788	r_{pb}^2 = 0.911
Differences in site	t stat = 3.228	t stat = -14.239	t stat = 12.864
(HUA site vs. LUA site)	t crit = +- 2.00	T crit = +- 1.98	t crit = +- 1.99
	r_{pb}^2 = 0.148	r_{pb}^2 = 0.659	r_{pb}^2 = 0.671
Differences in user groups	t stat = 0.075	t stat = -0.725	t stat = 0.455
(HUA users vs. LUA users)	t crit = +- 1.98	t crit = +-1.98	t crit = +- 1.98
Control for Usability	t stat = -0.553	t stat = -0.124	t stat = -0.117
(same dimension side for user and website vs. different dimension side)	t crit = +- 2.012	t crit = +- 2.012	t crit = +- 2.012

Table 4.22: Summary of Findings for Uncertainty Avoidance

It was noted that the differences found for the high uncertainty avoidant user group were exactly opposite to the differences found for the low uncertainty avoidant user group. This strongly suggested that the high uncertainty avoidant site was substantially superior to the low uncertainty avoidant site in terms of accuracy, speed and satisfaction levels. The independent samples t-tests confirmed that, irrespective of the user's side of the cultural dimension,

- higher levels of accuracy were achieved,
- less time was taken to complete the tasks, and
- greater satisfaction levels were reported,

when using the high uncertainty avoidant site than when using the low uncertainty avoidant site. This was confirmed by the lack of significant results found in the paired samples t-test used to control for usability. No significant difference was found between the two user groups.

4.5.3 Impact of Masculinity vs. Femininity on Usability

The only *insignificant* difference found at the 95% level in the comparisons was in the accuracy

scores between masculine users using a masculine site compared to the same users using a feminine site. This is reflected in Table 4.23.

Dimension Side	Accuracy	Speed	Satisfaction
MAS/MAS vs. MAS/FEM	t stat = 1.000	t stat = -3.878	t stat = 2.722
t crit = +- 2.11		r_{pb}^2 = 0.469	r_{pb}^2 = 0.304
FEM/FEM vs. FEM/MAS	t stat = -2.483	t stat = 6.645	t stat = -6.627
t crit = +- 2.024	r_{pb}^2 = 0.14	r_{pb}^2 = 0.537	r_{pb}^2 = 0.360
Differences in site	t stat = 2.599	t stat = -6.094	t stat = 4.497
(MAS site vs. FEM site)	t crit = +- 1.989	T crit = +- 1.981	t crit = +- 1.982
	r_{pb}^2 = 0.076	r_{pb}^2 = 0.249	r_{pb}^2 = 0.161
Differences in user groups	t stat = 0.189	t stat = 1.054	t stat = 0.749
(MAS users vs. FEM users)	t crit = +- 1.996	t crit = +-1.992	t crit = +- 1.997
Control for Usability	t stat = 0.595	t stat = 0.328	t stat = -0.584
(same dimension side for user and website vs. different dimension side)	t crit = +- 2.030	t crit = +- 2.030	t crit = +- 2.030

Table 4.23: Summary of Findings for Masculinity vs. Femininity

It was noted that the significant results found for the masculine user group were exactly opposite to the differences found for the feminine user group. This strongly suggested that the masculine site was substantially superior to the feminine site in terms of accuracy, speed and satisfaction levels. The independent samples t-tests confirmed that, irrespective of the user's side of the cultural dimension,

- higher levels of accuracy were achieved,
- less time was taken to complete the tasks, and
- greater satisfaction levels were reported,

when using the masculine site than when using the feminine site. This was confirmed by the lack of significant results found in the paired samples t-test used to control for usability. No significant difference was found between the two user groups

4.5.4 Impact of Individualism vs. Collectivism on Usability

There were no significant differences in the accuracy or speed levels achieved. Significant differences in user satisfaction occurred within the individualist user group, but not within the collectivist user group. These findings are reflected in table 4.24.

It was noted that the difference within the individualist user group was negative, indicating that individualist users using the collectivist site reported higher satisfaction levels than when using the individualist site. The independent samples t-test indicated that irrespective of the user's side of the cultural dimension, greater satisfaction levels were achieved overall when using the collectivist site than when using the individualist site. This was confirmed by the lack of significant results found in the paired samples t-test used to control for usability. No significant difference was found between the two user groups.

Dimension Side	Accuracy	Speed	Satisfaction
IND/IND vs. IND/COL t crit = +-2.03	t stat = -1.946	t stat = 0.220	t stat = -3.501 r_{pb}^2 = 0.254
COL/COL vs. COL/IND t crit = +-2.12	t stat = 1.376	t stat = 0.417	t stat = 1.029
Differences in site (COL site vs. IND site)	N/A	N/A	t stat = 3.562 t crit = +- 1.983 r_{pb}^2 = 0.108
Differences in user groups (COL users vs. IND users)	N/A	N/A	t stat = 0.265 t crit = +-1.996
Control for Usability (same dimension side for user and website vs. different dimension side)	N/A	t stat = 0.249 t crit = +- 2.030	N/A

Table 4.24: Summary of Findings for Individualism vs. Collectivism

4.6 Analysis and Interpretation

For every significant result obtained from the paired samples t-tests, a significant result was obtained from the independent samples t-tests for site usability differences. This indicated that the increase in performance could be attributable to the cultural dimension of the site, or variables other than that cultural dimension. In addition, the paired samples t-tests used to test site usability equality produced insignificant results, thus confirming the independent samples t-tests results in each case. This is reflected in Table 4.25.

Tests	Accuracy	Speed	Satis-faction	Accuracy	Speed	Satis-faction
	Power Distance			**Uncertainty Avoidance**		
Paired samples	Insignificant	Significant	Insignificant	Significant	Significant	Significant
Site Differences		Significant		Significant	Significant	Significant
User Differences		Insignificant		Insignificant	Insignificant	Insignificant
Usability Equality		Insignificant		Insignificant	Insignificant	Insignificant
	Masculinity vs. Femininity			**Individualism vs. Collectivism**		
Paired samples	Significant	Significant	Significant	Insignificant	Insignificant	Significant
Site Differences	Significant	Significant	Significant			Significant
User Differences	Insignificant	Insignificant	Insignificant			Insignificant
Usability Equality	Insignificant	Insignificant	Insignificant			Insignificant

Table 4.25: Summary of Results

The results of the statistical tests indicate that the differences in performance measures are attributable to variables in the test interfaces, rather than as a result of the user using an interface with a corresponding side of a cultural dimension. Consequently, there is insufficient evidence to support the hypotheses that any of the four cultural dimensions tested significantly affect user performance. To avoid a Type I error, all of the secondary null hypotheses were retained. This is depicted in Table 4.26.

The rejection of the secondary hypotheses at the 95% level resulted in the primary hypotheses (H1 – H4) being rejected, as reflected in Table 4.27. H5 could not be tested due to the limited number of test subjects that were identified as short-term oriented.

4.7 Lessons Learned and Further Work

Although significant differences were found in the paired samples t-tests, significant differences were also found between the sites. It is worth noting that the size effects of the site differences were between 50 – 80 percent of the size effect of the paired samples differences.

These results strongly suggest that the differences in user performance were attributable to one of the two sites in each set being a generally better site, rather than as a result of the cultural dimensions of the users. Some of the variables that could have contributed to this are suggested below.

Cultural Dimension	TEST	ACCURACY	SPEED	SATISFACTION
Power Distance (H1)	HPD/HPD vs. HPD/LPD	$H1_0A^1$ Retained	$H1_0T^1$ Retained	$H1_0S^1$ Retained
	LPD/LPD vs. LPD/HPD	$H1_0A^2$ Retained	$H1_0T^2$ Retained	$H1_0S^2$ Retained
Uncertainty Avoidance (H2)	HUA/HUA vs. HUA/LUA	$H2_0A^1$ Retained	$H2_0T^1$ Retained	$H2_0S^1$ Retained
	LUA/LUA vs LUA/HUA	$H2_0A^2$ Retained	$H2_0T^2$ Retained	$H2_0S^2$ Retained
Masculinity vs Femininity (H3)	MAS/MAS vs. MAS/FEM	$H3_0A^1$ Retained	$H3_0T^1$ Retained	$H3_0S^1$ Retained
	FEM/FEM vs. FEM/MAS	$H3_0A^2$ Retained	$H3_0T^2$ Retained	$H3_0S^2$ Retained
Individualism vs. Collectivism (H4)	IND/IND vs. IND/COL	$H4_0A^1$ Retained	$H4_0T^1$ Retained	$H4_0S^1$ Retained
	COL/COL vs COL/IND	$H4_0A^2$ Retained	$H4_0T^2$ Retained	$H4_0S^2$ Retained

Table 4.26: Summary of Hypotheses

	Hypothesis	Result
H1	Power Distance will affect the usability of a computer-based system.	$H1_0$ RETAINED
H2	Uncertainty Avoidance will affect the usability of a computer-based system	$H2_0$ RETAINED
H3	Masculinity vs. Femininity will affect the usability of a computer-based system	$H3_0$ RETAINED
H4	Individualism vs. Collectivism will affect the usability of a computer-based system	$H4_0$ RETAINED
H5	Short-term vs. Long-term Orientation will affect the usability of a computer-based system	Unknown

Table 4.27: Summary of Primary Hypotheses

- Nature of the cultural dimension: it is possible that one of the two sites in each set was a better site because it displayed the characteristics of a particular side of a cultural dimension.
- Usability characteristics: the better interfaces could have been more generally usable if the design of the interfaces encapsulated relevant usability principles, heuristics and guidelines.

Further analysis of the results indicated that the differences in scores could have been attributable to user variables other than the tested cultural dimensions. Moreover, additional unforeseen variables could have affected the cultural dimensions being tested. These variables could include:

- Cultural dimension strengths: it is possible that usability levels are significantly affected only if the users display high levels of a particular side of a cultural dimension. This aspect was not controlled for in the experiment, as it was noted that the majority of users scored in the low to medium levels of each dimension side. However, if the dimension sides themselves were inaccurately evaluated, it is equally possible that the strengths of the dimensions sides were also inaccurately evaluated.
- Interplay between dimensions: one cultural dimension could override the impact of the other cultural dimensions, particularly if the user displays a substantially high level of one dimension. This could result in some dimensions having little or no effect on the usability of the product.
- Relative impact of cultural dimensions: it is possible that only some of the cultural dimensions may have an impact on usability. As noted from the findings of the Smith and Chang [2003] study discussed in section 2.6.2.1, this also indicates the possibility that the different cultural dimensions may have a stronger or weaker effect on usability.
- Partial representation: it is possible that not all of the pages on each website displayed the side of the dimension being tested [Marcus, 2002]. This could have resulted in the user responding to different sides of the same dimension, thus distorting the results
- Other subjective cultural dimensions: subjective cultural dimensions other than those identified by Hofstede could have influenced the results. These other dimensions could have overridden the effects of Hofstede's dimensions on the users' performance.
- Performance determinants: although some of the more obvious determinants were controlled for, such as age, home language and computer literacy skills, it is possible that other objective cultural aspects and user characteristics could have influenced performance.

In addition to user and interface variables, the design of the experiment could have brought in additional confounding variables. Some of these are discussed below.

- Compulsory participation: participation in the experiment by the test subjects was compulsory. It is therefore possible that the students who did not want to participate performed at lower levels than usual.
- Time limit: a time limit was imposed on the task. This could have adversely affected the subjects' performance as it could have caused additional stress and anxiety.
- Adaptation of cultural profile questionnaire: the test subjects were students who have little or no working experience. Hofstede's cultural questionnaire was adapted to suit the context of the test subjects. As a result, the test subjects' cultural dimension sides could have been inaccurately evaluated.
- Order effect: of the four sets of sites tested for differences, the second site in three of the four sets appeared to have a higher usability than the first site. This suggests that the order in which the users actually used the interfaces could also have affected their performance levels.
- Component impact: usability could be affected at different levels by the various components comprising an interface. The results could have been affected if the test tasks did not test these components equally between the interfaces.
- Usability measures: usability was measured in terms of accuracy, speed and satisfaction. If a significance difference had been found for any one of these measures, the hypothesis would have been accepted. However, users with different cultural profiles may rank the importance of these measures differently. In this situation, the measures should be weighted according to their importance to establish the overall usability of the product

4.8 Summary

An experiment was undertaken to determine whether or not Hofstede's [2001] cultural dimensions should be accommodated into the design of user interfaces. This was done by testing the impact of these five cultural dimensions on the usability of computer-based interfaces. The accuracy, speed and performance levels of users were measured, using interfaces that corresponded to their side of each cultural dimension and comparing them to the same measurements taken whilst using interfaces with opposing sides of each dimension. For each side of each cultural dimension, a paired samples t-test was performed to determine whether the differences in mean scores were significant.

For every significant result obtained in the paired samples t-tests, independent samples t-tests were performed. In every case, significant differences were found between the two interfaces in each set of interfaces. These findings were confirmed by performing additional paired samples t-tests that compared accuracy, speed and satisfaction levels between users using the interface displaying one side of a cultural dimension, and users using the interface with the opposing side of the same cultural dimension. No significant differences were found in these tests.

The results of the experiment did not provide empirical evidence that a relationship exists between subjective culture and usability. This is in contrast to the theoretical evidence presented in Chapter 2 that such a relationship does exist. The statistical tests indicated that the results obtained from the experiment could have been influenced by variables other than the cultural dimensions being tested. This suggests that a more detailed and robust model of the variables that influence usability needs to be built before empirical research of this nature can be effectively conducted. A further investigation of the literature was conducted to identify the validity of these variables, the results of which are discussed in Chapter 5.