

Working Paper Series
ISSN 1170-487X

**Usability Testing:
A Malaysian Study**

**by Alvin Yeo, Robert Barbour
& Mark Apperley**

Working Paper 97/21
July 1997

© 1997 Robert Barbour and Alvin Yeo
Department of Computer Science
The University of Waikato
Private Bag 3105
Hamilton, New Zealand

Usability Testing: A Malaysian Study

Alvin Yeo

Computer Science Dept.
University of Waikato,
Private Bag 3105
Hamilton, New Zealand
awy@waikato.ac.nz

Robert Barbour

Science, Mathematics & Technology
Education Research Centre
University of Waikato,
Hamilton, New Zealand
r.barbour@waikato.ac.nz

Mark Apperley

Computer Science Dept.
University of Waikato,
Private Bag 3105
Hamilton, New Zealand
m.apperley@waikato.ac.nz

ABSTRACT

An exploratory study of software assessment techniques is conducted in Malaysia. Subjects in the study comprised staff members of a Malaysian university with a high Information Technology (IT) presence. The subjects assessed a spreadsheet tool with a Bahasa Melayu (Malaysia's national language) interface. Software evaluation techniques used include the think aloud method, interviews and the System Usability Scale. The responses in the various techniques used are reported and initial results indicate idiosyncratic behaviour of Malaysian subjects. The implications of the findings are also discussed.

Keywords

usability testing, think aloud, culture influence, power distance, Malaysia

INTRODUCTION

Today, US-based software developers earn more than half of their revenue from outside the United States (US). According to Fortune magazine, Microsoft derived 74% of its US\$11.4 billion revenue from outside US and Canadian markets in its fiscal year ending June 30, 1997 [1]. Furthermore, developing economies in Latin American, Eastern Europe, Africa and Asia have become Microsoft's fastest growing market earning about 10% of Microsoft's sales. The Asian growth trend confirms with the press release of Software Publishers Associations [2] which announced that the Asia Pacific region accounted for US\$1.4 billion of US personal application software sales in 1996, an increase of 22% over 1995. The fastest sales growth (over the 1995) were reported in China 81%, Malaysia 42%, Japan 27%, Thailand 24% and Korea 13%. Further increase in these sales are expected to continue.

With this growth potential, US software companies are also beginning to realise the need to improve their software to ensure a competitive advantage in the emerging markets. A product that is localised to its target market will more likely do well than one which is not localised. Furthermore, many companies have fallen into the trap that if a product is tested domestically and is doing well in the home market, it is believed that the product will do well in other countries [3]. One recent example, a sporting goods company had to retract its promotional merchandise (worth US\$1 million) in an Asian country as there was sexual connotations associated with the literal translation of its slogan "Just do it". This situation could have been avoided if the company

had ensured that their products were accepted with the local counterparts of their target markets, they would have saved time and effort. Likewise, if a software product is to succeed, it is crucial to conduct studies to ensure acceptance and usability of the product in the target market. The more one knows about a country, the greater the probability of success. From a usability testing perspective, there is little, if any, literature about international studies that have been conducted in developing countries in Asia such as Malaysia. It is our hope that we can find out more about Malaysian users in our study. This paper adds knowledge about usability testing in one of the emerging markets in Asia.

As the study was carried out in Malaysia, we will give a brief description about Malaysia in the next section. We then detail the method used in this usability study. The results and analysis of the study will be described in the ensuing sections. In the Discussion section, we will discuss the findings and its implications.

MALAYSIA

Malaya gained her independence from Britain in August 1957 and was joined by Sarawak and Sabah to form Malaysia in September 1963. Malaysia has a population of about 21 million [4]. The Malays and other indigenous groups make up 59% of the population, the Chinese, 32%, and the Indians, 9%. Bahasa Melayu, the mother tongue of the Malays, is the official and national language. Besides Bahasa Melayu and English, Chinese dialects, Tamil and other ethnic languages are also prevalent. As Malaysia was under British rule, the use of English is quite widespread. In the business sector, English is the *lingua franca* whereas Bahasa Melayu is used in official and administrative matters.

From 1957 to 1967, Bahasa Melayu and English were the official languages. From 1967 onwards, Bahasa Melayu became the sole medium of instruction in secondary schools [5]. Six years of primary school level education is available in Bahasa Melayu, Chinese and Tamil. To prepare the Chinese and Tamil-educated students for the secondary schools, these students enter "transition" (Bahasa Melayu-intensive) classes. Bahasa Melayu and English are compulsory in both primary and secondary education. Computer studies are now being introduced into the secondary school curriculum. These courses are taught in Bahasa Melayu but are likely to use English software. Similarly, computer courses are taught in Bahasa Melayu but using software in English.

Malaysia is currently undergoing rapid growth. Its economic growth rate has been around 8% for the past ten years. The Malaysia government has also set out a strategic plan, Vision 2020, to guide Malaysia to achieve fully developed status by the year 2020. The Malaysian Government has also initiated the Multimedia Super Corridor (MSC), "a vehicle to attract world class technology-led companies to Malaysia, and to develop local industries." [6]. Commercial software development in Malaysia comprises mainly in-house development. there is hardly any locally developed software applications for the public. One reason is that Malaysians are English literate and thus are likely to use the popular software applications from the US. Also, there is a high rate of piracy in Malaysia, about 80% [7] though this is likely to change with the MSC initiative as well as the introduction of the harsher penalties for pirating software under the new Computer Crime Bill 1997 [8].

METHOD

The study is mainly exploratory in nature. We attempt to investigate the use of the software assessment techniques such as the think aloud method (described as probably the most valuable usability engineering method [9]), logging, questionnaires and interviews using Malaysian subjects. In the next section we will describe the tool used. The ensuing sections will give details of the subjects and how the experiment was conducted.

Tool Used

The spreadsheet that was used in the experiment (also known as our First Internationalised Research Spreadsheet Tool or FIRST) has been extended to accommodate any language whose alphabet uses the Extended ASCII characters. The internationalisation technique is detailed in [10]. Currently the spreadsheet is available in English, Māori, Bahasa Melayu and Iban, an ethnic group in Malaysia. FIRST was modified from a spreadsheet application known as TCALC which is bundled with Borland's Turbo C++ compiler. Permission had been obtained from Borland International to use and modify their code. The spreadsheet is DOS-based and has a Lotus 123-like interface. The spreadsheet does not support mouse-use and thus relies on arrow keys for navigation. The spreadsheet has also been modified to record all keystrokes during any session. The keystrokes are also timestamped. Information such as what each cell contains as well as the language of the current interaction is also recorded.

The spreadsheet FIRST allows data entry of text, values and formulas. Other features available include save, print, insert and delete row and columns, formula display, automatic re-calculation.

The spreadsheet version used in the experiment has a Bahasa Melayu interface. The original version, TCALC, was in English. The Bahasa Melayu translation of the spreadsheet was conducted by a Malaysian university student proficient in Bahasa Melayu and English. The translation was verified by one of the authors of this article who knew how to speak and write Bahasa Melayu.

The Bahasa Melayu terminology used in the user interface was obtained from the *Istilah Komputer*, a collection of computer terminology published by the custodian of Bahasa Melayu, Dewan Bahasa dan Pustaka (the Malaysian Institute of Language and Literature).

Subjects

All the subjects in the experiment were volunteers from a local university in Malaysia. The university had a high IT presence whereby email and internet facilities were available to all students and staff of the university. The subjects' occupation ranged from clerks to managers. This diversity was to allow for a representative perspective from the different levels of the organisation. All the subjects chosen were expert spreadsheet users. Their reported use ranged from at least one hour a week to four hours a day. Experienced spreadsheet users were used as we believe they would be able to transfer their knowledge of one spreadsheet to another.

All the subjects were Malaysians. There were six Malays and two non-Malays. There were three females and five males in the study. An attempt was also made to obtain some subjects that the experimenter was not acquainted with.

Experiment Proper

The experiment was conducted in a room at the faculty where the experimenter works as a tutor. During the experiment, only the experimenter and the subject were present during the experiment. On arrival, the subject was informed about the details of the experiment and asked to complete an "Informed Consent" form. The subject was then requested for information pertaining to their spreadsheet experience and spreadsheet use. This step is to ensure that the subjects are indeed expert users. After the initial interview, the subjects are then given the tasks they are to complete in the experiment. As the subjects were required to think aloud in the experiment, the subjects were also briefed about how to "think aloud" while completing the tasks. The subjects were allowed to choose the language they could think aloud in, that is, in Bahasa Melayu, English or both. All the subjects were assigned to complete the same tasks in a spreadsheet with a Bahasa Melayu interface. These tasks include entering data, creating formulas in the spreadsheet, increasing the width of a column, changing the cells to currency format and exiting from the spreadsheet. Every subject's think aloud session was tape recorded.

Immediately after the completion of the tasks, the subjects filled a System Usability Scale. The System Usability Scale is identified as a "quick and dirty" method to gauge the users' response to the interface's usability [11].

After filling in the questionnaire, the users were then asked three questions: their opinion of the spreadsheet they had just used; what improvements, if any, would they like to see in the spreadsheet; and if they were given a spreadsheet in Bahasa Melayu or English, which would they prefer? After answering these questions, the subjects

were debriefed. The subjects were also given the opportunity to question the experimenter about experiment. The subjects were also requested not to talk to anyone about the experiment as not all subjects had completed the experiment. In the next section, we will describe the results of the think aloud session, System Usability Scale and the three-question-interview.

RESULTS.

The logged data from the spreadsheet were processed and presented in a more readable format for ease of analysis . An example of “processed” logged data can be seen in Figure 1. The logged data provided the keystroke pressed and also the time the key was pressed. The first line in Figure 1 contains the following information about the keystroke: B (language the spreadsheet is in i.e. Bahasa Melayu), A1 (the cell in which the keystroke was recorded, _CR_ (the key being depressed is the Carriage return) and the time that has elapsed since the spreadsheet was started.

B	A1	_CR_	0 min. 7.637 sec.
B	A1	_DOWN_	0 min. 8.132 sec.
A1	Food		
B	A2	R	0 min. 17.143 sec.
B	A2	a	0 min. 17.308 sec.
B	A2	t	0 min. 17.473 sec.
B	A2	e	0 min. 17.637 sec.
B	A2	s	0 min. 17.857 sec.
B	A2	_CR_	0 min. 20.000 sec.
B	A2	_DOWN_	0 min. 21.374 sec.

Figure 1 : Example of Processed Logged Data.

The audio record from the think aloud component was completely transcribed. The processed logged data was of immense help, especially in making sense of the audio log during the transcription process. The two data sources allowed the experimenter to get a clearer picture of the subjects’ interaction in the spreadsheet.

Think Aloud Session

Overall, all the subjects employed a trial and error strategy to complete the tasks. This strategy allowed the subjects to determine what commands to use to complete the prescribed tasks. The subjects mainly relied on their spreadsheet knowledge to complete the tasks.

Of the tasks given, the subjects had few problems with two tasks: the data entry and exiting from the spreadsheet. However, the subjects had some difficulty with widening the columns, creating formulas and currency-formatting some cells. These tasks were problematic as they required the subject to go through two levels of menus. For example, to widen the columns, the user would press / to get the first menu-level. From this list of command, the user would select Lajur (column). The user would be presented with another list of commands, the second-level menu. From this list, Lebar (widen) would be chosen. On the other hand, to exit the spreadsheet, the subjects only has to select / and then Keluar (Exit) to quite the spreadsheet.

In the next section, we detail the results from the System Usability Scale

System Usability Scale

The purpose of the System Usability Scale is to gauge the users’ response to the interface. The System Usability Scale comprises ten Likert scale questions. Although the System Usability Scale is only ten-questions long, it correlates well with SUMI [12]. A University of Cork study placed SUS correlation reliability of 0.8588 with SUMI [11].

1. I think that I would like to use this system frequently.

Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
1	2	3	4	5

Figure 2 : Question one of the System Usability Scale

The System Usability Scale is scored from 1 to 5, 5 for the most favourable response for the interface, 1, the least. Refer to Figure 2 for an example question in System Usability Scale. As you can see in the Table 1, subjects S3, S7, S12 and S16 would like to use the spreadsheet frequently. Subjects S2, S14, S10 would not. Overall only S4 and S16 rated software highly. The other five were non-committal whereas S12 had the highest disapproval rate.

Subject	S2	S3	S7	S12	S4	S14	S16	S10
Question 1	2	4	4	4	3	2	5	1
Average for the 10 questions	2.7	3.1	3.1	1.9	3.9	2.9	4.8	2.9

Table 1 : System Usability Scale Results

Interview

In the interview, the subjects were asked three questions. The responses to these questions were recorded on audio tape and later transcribed. The responses the questions can be found below. (The text in sans serif font are speech from the subjects, the italicised text are the translation of the Bahasa Melayu words. Text added by experimenter for clarity are enclosed in [] , and < > indicates words that could not be deciphered during transcription).

The first question seeked the subject’s opinion of the spreadsheet they have just used.

Responses ranged from those “who liked it” to those who found it “too outdated”. Subject S16 was most enthusiastic about the spreadsheet.

“Bagus! Senang digunakan. Boleh dikomersialkan, saya rasa.” Good! Easy to use. I believe it can be commercialised. [S16]

Other positive comments include:

“...I think it’s okay, it’s more, more like Lotus. I think we better use this in [the department] there.” [S7] and

“...saya rasa ia lebih I believe it’s more... effective if we use this.” [S12]

On the other end of the continuum, subject S10 was critical of the spreadsheet. According to subject S10:

"... if you compare with the... current spreadsheets we have in the market, of course that one [FIRST] is very outdated."

Other critical comments, include: "not very user friendly" [S2], "old fashioned" [S4], "It wasn't that friendly I have to be frank with you." [S14]. Subject S3 was more subtle,

"...it's not that difficult to use, [though] you still have to go aar... looking for things lah, and have a trial and error." [S3]

The second question asked what improvements could be made to the spreadsheet. Three of the subjects mentioned problems understanding the Bahasa Melayu commands. The main features suggested for improvements include use of a graphical user interface (GUI) environment which supports the use of mouse, windows and icons. The subjects also suggested inclusion of features such as a Help facility, cut and paste, file and data compatibility, and functions such as average.

In the third question, the subjects are asked which spreadsheet version they would choose if they were given a choice of the software in either Bahasa Melayu or English?

Subject S10 replied that he did not have any preference. The deciding factor in subject S10's case would be the proficiency of his subordinates. If his subordinates are proficient in English, he would consider using the English software, and likewise for Bahasa Melayu. Subject S16 was the only subject that preferred the Bahasa Melayu. S16 preferred Bahasa Melayu as he believed he would be able to learn the software quicker as his mother tongue was Bahasa Melayu. All the other six subjects preferred the English version. Their reasons were that there might be problems understanding the computer terminology in Bahasa Melayu given the proliferation of English software and the scarcity of software and references in Bahasa Melayu.

ANALYSIS

The following sections consist of behaviour gleaned from the think aloud and interview results. The next section reports the language used in the think aloud and interview sessions. Usability problems are then described in the following section. Details of behaviour during the think aloud are also presented.

Language Use

The subjects were given the choice to use either or both English and Bahasa Melayu in the think aloud and the interview sessions.

In the think aloud session, all the eight subjects used English. The speech of some of the subjects was interspersed with Bahasa Melayu. This observation excludes reading out loud the different Bahasa Melayu commands in the interface. Subject S16 had the most speech interspersed with Bahasa Melayu whereas subjects S2, S4, S7, S10 and S12 had at most, two sentences in Bahasa Melayu. The use of Bahasa Melayu ranged from

the reference of cell addresses to comments of the spreadsheet. Examples of the Bahasa Melayu use include:

Trying out to figure out how to select the command:

Ah, macam mana? *How?* [S16]

Subject S16, commenting on the spreadsheet:

Ini sistem tak best ni. *This system is not "good".* [S16]

Subject S10, when trying to format cell B2 to B2:

"Umm, format, B dua, B dua, oh... okay. Eh? mana boleh... *Umm, format, B two, B two, oh... okay. Eh? how can...* [S10]

Subject S3 after managing to widen the cell width.

"...so I manage to get that aar... cell dilebarkan *widened*" [S3]

Subject S14 while expecting the spreadsheet to prompt her for a cell address:

I was expecting that tapi tak ada... *but did not appear...* [S14]

Subject S12, in navigating through the spreadsheet

Saya guna... *I use I use joystick [control keys] to remove... to remove...* [S12]

The language used in the interview is similar to the language used in the think aloud session. Some subjects using Bahasa Melayu interspersed with English whereas others do not use, if at all, Bahasa Melayu in their responses. Only Subject S16 spoke mainly in Bahasa Melayu in the interview. Subjects S3, S12 and S14 lapsed into Bahasa Melayu at times whereas subjects S2, S4, S7 and S10 did not use Bahasa Melayu other than to refer to command or components of the Bahasa Melayu user interface.

The following examples of Bahasa Melayu-use are replies to Question 3 of the interview i.e. "If provided software in English and Bahasa Melayu, which would you choose?":

For the time being, stick with English because kalau software dalam Bahasa Melayu, besar kemungkinan tak faham dengan istilah-istilah Bahasa Melayu. *For the time being, stick with English because if [we use] software in Bahasa Melayu, there is a high likelihood that we may not understand the Bahasa Melayu terminology.* [S3]

Ummm, I think English is more easier because it command is always... what we call... selari *in line with* when, when, when we use it... compared with Bahasa Melayu lah... [S12]

Ah... well, if you can have the English version in help... bahagian bantuan *help section*, then ... because I mean after all, if you want to come up with something in Bahasa Malaysia [Bahasa Melayu], at least they can refer to < > BI [Bahasa Ingeris *English*], English version. [S14]

I rasa sebab bila you guna Bahasa Melayu, you lagi 'ni, I rasa, lagi cepat, you... ni. Lagi cepat you tangkap, I rasa. *I feel that if you use Bahasa Melayu, you will pick up/learn [the software] much faster.* [S16].

In the next section, we will describe usability problems uncovered in the think aloud transcription.

Usability Problems

In this paper, we use Nielsen's [9] usability problem definition, "...a usability problem is any aspect of a user interface that is expected to cause users problems with respect to some salient usability measure and that can be attributed to a single design aspect."

A number of usability problems were discovered after examining the think aloud transcription. As we are unable to detail all the usability problems found, we have grouped the usability problems in the following categories. The usability problems encountered include:

1. Bahasa Melayu terminology

The subjects had problems with the Bahasa Melayu commands because the subjects did not understand or were unfamiliar with the terminology used. For example, four subjects initially did not understand the word Lajur or "Column" in the spreadsheet context. The subjects usually had to translate the Bahasa Melayu commands to English before they could understand the commands.

In another example, the users assumed the use of English instead of Bahasa Melayu in the software. In some situations, the users are required to answer prompts like "Do you want to overwrite the old file?". Subjects would need to enter either T for Tidak (No) or Y for Ya (Yes). However, five of the eight subjects answered Y (as in Yes) and N (as in No) instead of the correct Y for Ya and T for Tidak.

2. GUI preference

Subjects preferred a GUI interface. Subjects actively looked for the mouse. The subjects also had problems trying to adapt to the command-line user interface as opposed to GUI. The subjects also looked for "icons" (buttons) to do the SUM of cells.

3. FIRST as a spreadsheet

As the users were spreadsheet users, the subjects expected FIRST to work like the Excel spreadsheet they were used to. Among the "features" the subjects looked for were, filename to be displayed on the spreadsheet when the spreadsheet is saved. The subjects also searched for commands like undo, cut and save, and help which were not supported in the spreadsheet FIRST. One subject pressed F1 for help while another tried the key combination Ctrl-F to get the File menu. (Ctrl-F is the key combination to get to the File menu on software running under the Windows operating system.) One subject looked under the command Sunting (Edit) expecting to find the Bahasa Melayu-equivalent of commands Cut and Paste. (In most applications, the Cut and Paste commands do appear in the Edit menu).

The next section will report on behaviour observed in the think aloud session.

Behaviour in the Think Aloud Session

Prompting is kept to a minimum so as not to disrupt the subjects train of thought. However, in some situations whereby the subject falls silent for, say, more than a minute, the experimenter will ask the subject, "What are you thinking of now?". Otherwise, we may have little

commentary if hardly at all from the more quiet subjects. In our study, some users needed more prompting than others. Subject S7 had to be prompted to keep thinking aloud. However, despite the calls to think aloud, she remained quiet almost throughout the session. From the Table 2, we can see that subject S7 has the lowest number of words spoken per minute as compared to the other subjects.

On the other end of the spectrum, some subjects needed little encouragement to think aloud. Subjects like S4, S16 and S10 were quite free with their comments.

Subject	S2	S3	S7	S12	S4	S14	S16	S10
Word ¹ count	1493	964	291	1291	1395	979	1164	2098
Time to finish experiment (mins.)	36.7	32.9	18.5	31.9	22.1	23.8	15.3	29.7
Word/min	41	29	16	40	63	41	76	71

Table 2 : Words Spoken During Think Aloud Session

Negative Comments

Another observation made in the think aloud and interview session, was that certain subjects were quite frank with their comments. (The authors had hoped that the users would comment more on the spreadsheet especially that the spreadsheet FIRST was a very simple DOS-based spreadsheet which does pale in comparison with current spreadsheet offerings.) These comments also found in the second question in the interview.

I think the first time when I see the spreadsheet, I thought it was easy to use, but after trying so many thing doesn't work, then arr... ..I found out that the system is urr what we call... it's not that [bad] a utility to start [with] ... that it's not that difficult to use. You still have to go aar... looking for things [commands] lah, and have a trial and error. That's what I did. [S3-interview]

I feel, I feel, I... it's better if I use a window because this DOS is not really effective... I think so... [Laughs] [S12-talk aloud]

Teruk jugak spreadsheet tu. *This spreadsheet is "bad"*. [S2- interview]

I don't like your interface. [Laughs.] Too primitive. [S4]ta. Umm, old fashioned, umm, reminded me of Lotus 123. [S4-interview]

I was taken aback. It wasn't that friendly I have to be frank with you. [Laughs.]-[S14-interview]

Ini system tak best ni. *This system is not "good"*. [S16-talk aloud], when he wasn't sure whether to answer yes or no or Ya or tidak.

Of course, if you compare with the... current spreadsheets we have in the market, of course that one is very outdated. ...So this one is very basic lah, in a sense. [S10-interview]

¹ The "words" here included utterances such as filled pauses like ummms and ahhs.

Subject S7 did not make any direct negative comments about the spreadsheet.

We discuss the observations made in the following section.

DISCUSSION

As most Malaysians use both English and Bahasa Melayu, this Malaysian characteristic is observed in the think aloud and the interview sessions in which English is used interspersed with Bahasa Melayu, and vice versa. As described in the Language Use section, there are varying degrees of Bahasa Melayu-use. Thus, it would be advisable to have the experimenter conducting the experiment to be well versed in both Bahasa Melayu and English. Otherwise, the subject may have difficulty in expressing him or herself. It is usually difficult in describing something in a language that is not the mother tongue. Analysis of audio and video recordings may require the services of a Malaysian who is well versed in the languages to read between the lines. With voice recognition seen as the future of human-computer interaction, dictionaries of English and Bahasa Melayu may be required to enhance the effectiveness of voice recognition in Malaysia.

The person who is conducting the experiment may also affect the experiment. For example, the experimenter's Bahasa Melayu was not as good as his English, thus subject S12, whose first language is Bahasa Melayu decided to speak in English. This choice is evident in the think aloud session and interview whereby S12 (a clerk) began a sentence in Bahasa Melayu but completed the sentence in English. His speech was a bit hesitant. This "phenomenon" is similar to the observations made by Elaine Morais [13] of Universiti Malaya in which she describes the use of Malaysian English spoken by white collar and blue collar workers. "Top level management used English predominantly while throwing in a few words and phrases in Bahasa while blue collar workers (most of whom were Malay) spoke in the national language, interspersed with English." [13]. When both levels meet in a formalised environment, e.g. in a production meeting, a convergent strategy is used. Top management used more Malay to reduce the status gap and to reach out to their employees while the workers used more English. Thus, a lower ranked Malay may be more effective eliciting information from lower ranked users.

Another interesting point is that, although, the subjects are Malays whose mother tongue is Bahasa Melayu, they still preferred English software on the whole. This preference is probably due to the fact that the software they have been exposed to and have used is in English. However, subject S16 preferred software in Bahasa Melayu as he had no problems with the Bahasa Melayu terminology. (S16 is a native Bahasa Melayu speaker and also a computer science tutor.)

From the Usability Problems section, we can see that usability testing techniques such as think aloud are

effective in finding usability "catastrophes" in Malaysia and shows that Western techniques can work in an Eastern scenario. The usability problems in the Malaysian study found could be aggregated and compared with Nielsen's [13] study which he estimated the number of subjects needed for a thinking aloud test and found that using five subjects will uncover 77-85% of all usability problems.

Playing devil's advocate, we deliberately use a DOS spreadsheet with little functionality in the usability testing with the expectation that the users would criticise the software vehemently. However, from Table 3, only S10 was quite critical of the software indicated by the use of superlatives. Few people will criticise negatively as this may be attributed to the Malaysian behaviour *jaga maruah*, the preservation of face. Preservation of face means that one should maintain a person's dignity by not embarrassing him or communicating negative feedback and embarrassing him or her [15]. Furthermore, from the System Usability Scale results (see Table 3), the average scores were almost all either 3 or greater – a score of 1 is least favourable response to interface, score of 5 most favourable response, score of 3 is neutral. (S16's score does not count as he had no problems with the spreadsheet and he really liked the software). This neutral/positive response supports Herman's [16] observation of Singaporean users' subjective evaluation which tended towards the positive despite clear indications of poor user performance. Similarly in our study, despite having problems with the interface, a command-line DOS interface at that, the users still gave relatively positive response. Note also that three subjects gave a score of 4 (i.e somewhat agree) to question one of the System Usability Scale in which the users are asked if they would like to use the software frequently.

The observations above could be explained from the cultural perspective of power distance. Power distance is defined as "the extent to which the less powerful members of institutions and organisations with a country expect and accept power is distributed unequally." [17]. Malaysia has been identified by Hofstede as the country having the highest power distance. This means that Malaysians in general are willing to accept the fact that inequality in power is considered normal and thus their authority, unquestioned. Some of the attributes of a high power distance country are that power holders are entitled to material privileges such as a big office or big car. Also, employees are "afraid" of their employers as their employers wield powers such as the authority to fire employees. Relating back to our study, a person of higher status and power will be more likely to voice his or her feelings of discontent than one from the lower ranks. A higher ranked person would have less fear of retribution than one say, of a slightly lower rank. The Table 3 shows the order or hierarchy in the university. Comments from higher level staff tend to be more critical compared to staff of lower rank. From the System Usability Scale scores of question 1, scores of 4 appear in three of the lower ranked subjects.

Occupation	Gender	Familiar	Subject	Comments	SUS Q1	SUS Average
Manager	Male	No	S10	Very outdated, very basic	1	2.9
Lecturer	Male	Yes	S4	Too primitive, old fashioned	3	3.9
Manager	Female	Yes	S14	Not that friendly	2	2.9
Tutor (Experimenter)	Male	–	–	–	–	–
Tutor	Male	Yes	S16	This system not “good”.	5	4.8
Financial Assistant	Female	Yes	S2	“Bad” spreadsheet	2	2.7
Clerk	Male	No	S3	Still have to go looking for things	4	3.1
Clerk	Male	Yes	S12	Not really effective,	4	1.9
Data Processing Operator	Female	No	S7	No negative comment.	4	3.1

Table 3 : Comments and System Usability Scale Ordered by Ranking From High to Low Rank

Subject S12's score of 2 could be attributed to how he perceived his power relation with the experimenter, that he is not “afraid” of the experimenter. S7 was quite reticent in her think aloud, and this may be due to her being of a lower rank. However, we cannot claim that the behaviour is entirely attributed to her ranking and that it might be her personality that she is quiet. As we believe that someone of a higher rank would be more likely to make more critical or negative comments, their usability assessment may be more accurate.

Other factors that are likely to play a role in the study is the gender. Studies have shown that women tend to be less talkative than men. Another factor may be the familiarity with the subject. It would seem from the results below that if someone is familiar to the experimenter, a more frank response is more likely.

SUMMARY

Despite the small sample, there were some interesting findings in the study. For example, the use of English, Bahasa Melayu and Malaysian English in the think aloud and interview sessions. The results justify the use of local usability testers in light of the language use of the users. Although the think aloud and interview techniques are originate from the West, there is indication that these techniques would work provided the following factors were accounted for: rank of the subjects, gender and the familiarity of the experimenter with the subject. These factors should be tested in countries which share the same cultural background as Malaysians. As with most qualitative results, these results are indicative and further work looking at a greater number of subjects is recommended.

ACKNOWLEDGMENTS

The authors would like to thank Borland International for permission to use the TCALC source code in our research.

REFERENCES

1. Brent Schlender. Microsoft: First America, Now the World. *Fortune*. 18 Aug 1997. Available as http://www.pathfinder.com/@@floq*wcA2*ukZ3eq/fortune/1997/970818/fea.html.
2. Software Publishers Association: Asia Pacific Application Software Sales Rise 22% in 1996. Available as <http://www.spa.org/research/releases/96q4asia.htm>.
3. Fernandes, T. *Global Interface Design*. Academic Press, Chestnut Hill, MA, 1995.
4. Department of Statistics. *Yearbook of Statistics (Malaysia) 1996*, Department of Statistics, Malaysia. 1996.
5. Pakir, A. Making Bilingualism Work: Development in Bilingual Education in Asean, In Khoo, R, Kreer, U., And Wong, R, (eds.) *Towards global Multilingualism: European Models and Asian Realities*, Multilingual Matters Ltd., Clevedon, England, 13-27, 1994.
6. Multimedia Super Corridor 1997. Available as <http://www.mdc.com.my/msc/index.html>.
7. NST. Asia-Pacific Suffers Worst Software Piracy Losses. Business Section. *New Straits Times Newspaper*, 23, 9 May 1997 issue, 1997.
8. Computer Crime Bill 1997. Available as <http://thestar.com.my/archives/infolib/compcrime.html>
9. Nielsen, J. *Usability Engineering*. Paperback Edition. Academic Press, 1993.
10. Barbour, R., and Yeo, A. Internationalising a Spreadsheet for Pacific Languages. *New Zealand Journal of Computing*. (Forthcoming). 1996.
11. Holyer, A. Methods for Evaluating User Interfaces. *Cognitive Science Research Paper No. 301, School of Cognitive and Computing Sciences, University of*

- Sussex*. Available as <http://www.cogs.susx.ac.uk/cgi-bin/htmlcogsreps?csrp301>.
12. Brooke, J. B. SUS - A Quick and Dirty Usability Scale . In Jordan, P (ed.) *Usability Evaluation in Industry*. Taylor and Francis, 1995.
 13. Morais, E. Talking in English But Thinking Like a Malaysian: Insights from a Car Assembly Plant. In *Proceedings of English Is An Asian Language: The Malaysian Context Symposium*, (Kuala Lumpur, Malaysia, August 1997).
 14. Nielsen, J. Estimating the Number of Subjects Needed for a Thinking Aloud Test. *International Journal of Human-Computer Studies*, 41, 385-397, 1994.
 15. Abdullah, Asma. *Going Glocal*. Malaysian Institute of Management. SNP Offset, Shah Alam, Malaysia, 1996.
 16. Herman, L. Towards Effective Usability Evaluation in Asia: Cross-cultural Differences. In Grundy, J. and Apperley, M. (eds.), *Proceedings of OZCHI'96: Sixth Australasian Computer Human Interaction Conference*. (Hamilton, New Zealand, Nov., 1996), IEEE 135-136.
 17. Hofstede, G. *Cultures and Organizations: Software of the Mind*. Paperback Edition. HapperCollins, Glasgow, 1994.