EVALUATION OF SOFTWARE USED IN AN INTERNET-BASED SIMULATION: ARE THERE ANY CULTURAL DIFFERENCES?

Knut Ekker
Arnstein Eidsmo
Nord-Trøndelag University College
Box 2501, N-7729 Steinkjer, Norway

ABSTRACT

This paper presents an evaluation of the software used in an Internet-based simulation / role playing environment with a focus on identifying cross-cultural differences in experience with the simulation / role playing software. The research model also identifies to what extent the individual characteristics of the participants also may influence the evaluation of the software used in the Internet-based simulations.

The individual characteristics focused on in this study (sex, age and computer knowledge) do not have an impact on the overall reaction to the software used. Similarly, the cultural differences between German and Spanish students do not generate different evaluations of the software used. There are some effects on the more specific evaluations of the software, however. Gender, major in college and computer literacy have an impact on the evaluation of the system capabilities of the software. Female students, students in the humanities and social sciences, and students with a low computer literacy are more likely to have a positive evaluation of the system capabilities. The user evaluation of the terminology of the system and the evaluation of the system capabilities do have an effect on the overall reaction to the software in that the more positive the evaluation of the former, the more positive is the overall reaction.

KEYWORDS

Usability, human factors, cultural differences, simulation, game, role-playing

1. INTRODUCTION

The research on which this paper is based is an ongoing data collection effort initialized by the IDEELS project – Intercultural Dynamics in European Education through onLine Simulation – a project funded by the Socrates/Erasmus Program of the European Union. The IDEELS project utilized a software package developed at the University of Bremen – OPUSi – a textbased conference environment that allowed interaction between student participants in simulations / role-playing games. The interaction between the students utilized both symmetric and asymmetric communication such as internal emails, memos and "live" conferences (Sutherland, 2003).

This paper presents an evaluation of the software used by the participants in the IDEELS simulation from 1998 until 2004. The research model employs user satisfaction scales developed in the last half of the 1980's by Ben Schneiderman (1987) and refined by Chin et al (1988). Focus on users satisfaction has continued through the 1990's and into this century. User satisfaction has been a central element of the evaluation of human-computer interaction (Dix, 2004).

Researchers have also focused on the attitudes of users as an indicator / predictor of the users eventual acceptance of the software applications (Joshi, 1992; Davis, 1993). The link between attitudes and behavior has been outlined in the Theory of Reasoned Action by Ajzen and Fishbein (1980), identifying a causal link between beliefs, attitudes and behavior. Over the last decade there has been several research projects focusing on the predictors of attitudes towards use of computers and software, and a set of variables - gender, age, personality type, degree of expertise and computer literacy - are central in predicting attitudes towards computers and end-user satisfaction. (Prince, 2004; Orr, 2003; Houle, 1996).

The focus on this paper is the analysis of factors that may predict how different groups of participants evaluate the software used in the internet based simulation.

A model of the research question is presented below:

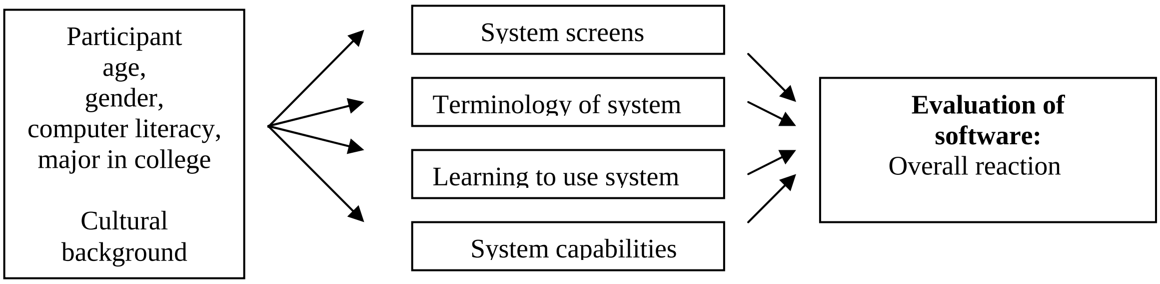

Figure 1. Evaluation of software, cultural differences and participant characteristics

2. ANALYSIS

Students from several countries in Europe participated in the IDEELS simulations over the Internet. The student sample is a heterogeneous group studying information technology, English as a second language, engineering and social sciences. The participants in the IDEELS simulations act as negotiators, technical consultants, activists or journalists within the “Eutropian Federation”. During the course of three-week long simulations, the participants communicate by exchanging messages, papers and communicating via textbased conference situations. The software used was a web-based interface driven by a database server (Sutherland, 2003).

In order to evaluate the software used in this simulation / role playing project, a web-based questionnaire was used. The survey included a set of questions measuring user satisfaction (Chin, 1988) in addition to background characteristics and cultural differences measured by a focus on students in Northern vs. Southern Europe (German vs Spanish as the native language).

2.1 Background characteristics of participants

Cultural background will be used as an explanatary variable in this analysis. The IDEELS project has had participants from all over Europe and in some cases from the Middle East. In order to focus on different cultural background, this analysis will focus on German students (University of Bremen) versus Spanish students (Polytechnic University of Valencia).

The background characteristics of participants are shown in table 1. About half of the student sample were 22 years of age or younger. The self assessment of computer knowledge indicated that about 55 % of the students regarded themselves as “Very good” or “Excellent”.

Table 1. Background characteristics of participants. Percentages. N=190

Variable		
Gender	Male	68,4
	Female	31,6
Age	20 and under	4,3
	21 – 22	43,6
	23 – 24	34,6
	25 and over	17,6
Computer knowledge	Poor	9,0

	Good	35,6
	Very good	35,1
	Excellent	20,2
Major in college	Hum. / soc. sci.	21,6
	IT / engineering	78,4
Cultural background (native language)	German	26,8
	Spanish	73,2

2.2 Simulation topics

The topics that the participants grappled with during these simulations were: educational system design and finance (1998), tertiary education (1999), human rights (2000), information and communication technology issues (2001, 2004) and migration (2003). There were no three-week simulation in 2002. Although the topics of the simulations varied widely, the roles of the students were quite comparable across all four simulations. The students acted as negotiators, consultants, journalists and activists building coalitions, arguing their positions and to varying degrees achieving consensus and ratification of an agreement on the topic of the simulation at the end of the three-week period. The experience of the simulation was therefore quite similar for all participants even though the topic of the simulations varied.

2.3 User satisfaction scales: overall reaction and four specific evaluations

The distribution of item averages and standard deviations for the user satisfaction measures are shown in tables 2 and 3. These measures were developed in the late 1980's (Shneiderman, 1987; Chin et al, 1988) and the items have high face validity.

Each item in this survey of users was measured with a value range of 5 points – not the 11 points used in the original scale developed by Shneiderman (1987) nor the 10 point value range presented by Chin (1988).

Table 2 shows the overall reaction to and the specific evaluations of the software. With reliability scores (Cronbach's alpha) of 0,79 and 0,68, the 6 items in the "overall" scale were all included. The "Difficult – Easy" item in the "overall" scale could have been removed in order to attain a higher value of alpha, but it was decided to keep the scale with all six items in order to make it more comparable with the original scale.

All items in the four specific evaluation scales are included, since the Cronbach alpha (scale alpha) is between 0,68 and 0,72.

Table 2.User satisfaction: Overall reaction & specific evaluations (3-week simulations 1998-2004)

User satisfaction – overall reaction			Mean (st.dev)	Cronbach alpha if item deleted
Items: overall reaction to the software				
Terrible	-	Wonderful	3,1 (0,88)	0,74
Difficult	-	Easy	4,0 (0,92)	0,82
Frustrating	-	Satisfying	3,1 (1,02)	0,74
Inadequate power	-	Adequate power	3,1 (1,00)	0,75
Dull	-	Stimulating	3,1 (0,76)	0,75
Rigid	-	Flexible	3,1 (0,97)	0,75
Overall reaction scale: sum of 6 items / (N of variables)			Scale alpha = 0,79	N=165

User satisfaction – screen evaluation				Mean (st.dev)	Cronbach alpha if item deleted
Items: Screen					
Reading characters	Hard	-	Easy	3,4 (1,16)	0,71
Highlighting simplifies task	Not at all	-	Very much	3,3 (0,81)	0,62
Organization of information	Confusing	-	Very clear	3,4 (1,01)	0,55
Sequence of screens	Confusing	-	Very clear	3,3 (1,04)	0,56
Screen evaluation scale: sum of 4 items / (N of variables)				Scale alpha = 0,68	N=176

User satisfaction – Terminology and system information				Mean (st.dev)	Cronbach alpha if item deleted
Items: Terminology and system information					
Use of terms throughout system	Inconsistent	-	Consistent	3,5 (0,75)	0,67
Terminology related to task	Never	-	Always	3,5 (0,70)	0,66
Position of messages on screen	Inconsistent	-	Consistent	3,4 (0,90)	0,64
Prompts for input	Confusing	-	Clear	3,3 (0,92)	0,61
Computer informs about its progress	Never	-	Always	3,0 (0,87)	0,62
Error messages	Unhelpful	-	Helpful	3,1 (1,08)	0,70
Terminology scale: sum of 6 items / (N of variables)			Scale alpha = 0,72		N=138

User satisfaction – Learning to use the system				Mean (st.dev)	Cronbach alpha if item deleted
Items: Learning					
Learning to operate the system	Difficult	-	Easy	4,1 (0,85)	0,66
Exploring new features by trial and error	Difficult	-	Easy	3,6 (0,89)	0,61
Remembering names and use of commands	Difficult	-	Easy	3,8 (0,89)	0,63
Performing tasks is straightforward	Never	-	Always	3,4 (0,72)	0,65
Help messages on the screen	Unhelpful	-	Helpful	3,3 (1,00)	0,67
Supplemental reference materials	Confusing	-	Clear	3,3 (0,95)	0,68
Learning scale: sum of 6 items / (N of variables)	Scale alpha = 0,69				N=138

User satisfaction – System capabilities				Mean (st.dev)	Cronbach alpha if item deleted
Items: System capabilities					
System speed	too slow	-	fast enough	2,8 (1,28)	0,60
System reliability	unreliable	-	reliable	2,8 (1,07)	0,59
System tends to be ...	noisy	-	quiet	3,3 (1,02)	0,66
Correcting your mistakes ...	difficult	-	easy	3,2 (0,89)	0,67
Designed for all level of users	never	-	always	3,6 (0,89)	0,71
System capabilities scale: sum of 5 items / (N of variables)			Scale alpha = 0,70		N=148

The distributions of all five scales are close to the normal distribution, and the four specific evaluations will be used as intervening variables in the regression analysis with the overall reaction to the software as the dependent variable.

2.4 Overall reaction to the software and specific user satisfaction variables.

Table 3 presents the results from the regression model analyzing the multivariate relationships between the intervening variables (four specific evaluation scales) and the overall reaction to the software as the dependent variable.

Table 3. Overall reaction to software and intervening variables measuring user satisfaction. N=105

Variable in regression Dependent variable: Overall reaction to software	Unstandardized Coefficients B	 Std.error	Standardized coefficients Beta	 t	 Sig.
(Constant)	0,37	0,31		1,20	0,23
System screens	0,04	0,08	0,05	0,52	0,60
Terminology of system	**0,38**	0,11	0,34	3,34	**0,001**
Learning to use system	0,08	0,10	0,07	0,84	0,40
System capabilities	**0,39**	0,07	0,45	5,71	**<0,001**
R = 0,74 R^2 = 0,55 (adjusted R^2 = 0,53)					

Two of the four intervening variables evaluating the specific aspects of the software contribute significantly towards explaining the overall reaction to the system. The evaluation of the terminology of the system and the evaluation of the system capabilities both contribute towards explaining as much as 55% of the variation in the overall reaction to the software of the simulation.

The four intervening variables measure aspects of the user satisfaction, and the question of a problem with multicollinearity arises. The bivariate correlations between these four variables are in the range from r=0,34 to r=0,65 – and the correlation of these four variables with the dependent variable of the overall reaction to the software is between r=0,50 to r=0,65. The collinearity statistics show an acceptable tolerance of 0,45 (VIF=2,24) of the terminology of system variable and 0,74 for the system capability variable. The two excluded variables of system screens and learning to use the system also show acceptable tolerance values of 0,51 and 0,62 respectively. When the tolerence value is less than 0,25 (VIF > 4.0) there is a problem with multicollinearity to the extent that the standard error (SE) is inflated (doubled with a tolerance of 0,25).

Table 4 and 5 use analysis of variance (ANOVA) in order to examine the effects of the background variables (gender, computer knowledge, major in college and cultural background – German vs. Spanish as their native language) on the four intervening variables.

There are only a couple of significant (or near significant) effects in tables 5 and 6, and they are highlighted in the table. Gender and major in college both have an effect on the evaluation of system capabilities. Female students are more likely to give a positive evaluation of the system capabilities (average score for female students is 3,34 vs 3,07 for male students; F=4,35, sig.=0,04). The data analysis show no difference between the German and the Spanish participants.

There is a also a small difference with respect to major in college. The students in the humanities and social sciences have a slightly higher score (more satisfied with the system capabilities) compared with the IT and engineering students (average score of 3,4 vs. 3,10; F=3,54, sig.=0,06).

Computer literacy shows a similar minor effect – students with low computer literacy give a more positive evaluation of the system capabilities compared with students with a high computer literacy score (3,3 vs. 3,1; F=3,57, sig.=0,06). Age does not have an effect on any of the intervening variables. These results are not included in the tables.

Table 4. Evaluation of system screens and terminology of system and characteristics of participants.

Characteristic of participants		**System screens**			**Terminology of system**		
		Mean	Std. Error	F-test	Mean	Std. Error	F-test
Gender	Male	3,31	0,07	F=0,06	3,29	0,06	F=0,12
	Female	3,34	0,09	sig.=0,81	3,33	0,08	sig.= 0,73
	Total	3,32	0,05	N=176	3,30	0,05	N=138
Major in college	Hum./soc.sci	3,40	0,11	F=0,54	3,37	0,11	F=0,40
	IT/engineering	3,30	0,06	sig.=0,46	3,29	0,05	sig.=0,53
	Total	3,32	0,05	N=176	3,30	0,05	N=138
Cultural background	German	3,26	0,11	F=0,32	3,20	0,12	F=1,14
	Spanish	3,33	0,06	sig.=0,57	3,33	0,05	sig.=0,29
	Total	3,32	0,05	N=176	3,30	0,05	N=138

Table 5. Learning to use system and system capabilities and characteristics of participants.

Characteristic of participants		**Learning to use system**			**System capabilities**		
		Mean	Std. Error	F-test	Mean	Std. Error	F-test
Gender	Male	3,57	0,06	F=0,04	**3,07**	**0,07**	**F=4,35**
	Female	3,60	0,08	sig.=0,84	**3,34**	**0,09**	**sig.= 0,04**
	Total	3,58	0,05	N=138	3,15	0,06	N=148
Major in College	Hum./soc.sci	3,64	0,11	F=0,31	**3,40**	**0,11**	**F=3,54**
	IT/engineering	3,57	0,05	sig.=0,58	**3,10**	**0,06**	**sig.=0,06**
	Total	3,58	0,05	N=138	3,15	0,06	N=148
Cultural background	German	3,51	0,10	F=0,61	3,17	0,13	F=0,02
	Spanish	3,60	0,05	sig.=0,43	3,15	0,07	sig.=0,88
	Total	3,58	0,05	N=138	3,15	0,06	N=148

Figure 2 shows that there is no interaction effect between gender and major in college on evaluation of system capabilities (ANOVA interaction results: F=0,20, sig. = 0,66). The figure does show, however, the gender effect clearly where female students are more satisfied with the system capabilities (average of 3,34) compared with male students with an average score of 3,07 as well as the effect of major in college (3,4 vs 3,1). These averages reflect that there are more female students in the humanities and social sciences (68%) and a majority of male students in the IT and engineering field (79%).

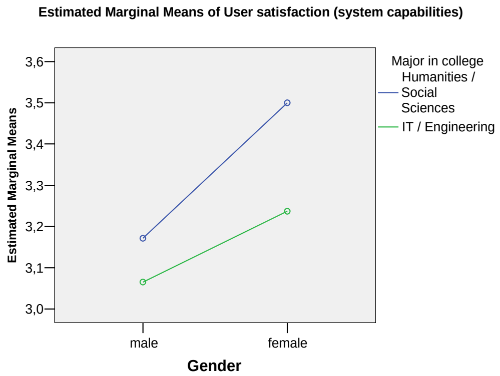

Figure 2. System capabilities, gender and major in college: analysis for interaction effects.

2.5 Evaluation of the software used and background characteristics

Table 6 shows the results of the multiple regression of the background characteristics (dummy coded variables), the specific evaluation of the software (intervening variables) and the overall reaction to the software used in the simulation. The only significant effects are the effects of two of the intervening variables (terminology of the system and system capabilities) on the overall reaction to the system software. None of the background characteristics contribute towards explaining variation in the overall evaluation of the software used in the simulation.

Table 6. Evaluation of the software used and background characteristics of participants. N=103

Variable in regression Dependent variable: Overall reaction to software	Unstandardized coefficients B	 Std.error	Standardized coefficients Beta	t	Sig.
(Constant)	0,55	0,66		0,85	0,40
System screens	0,05	0,09	0,05	0,54	0,59
Terminology of system	**0,34**	0,12	0,30	2,80	**0,006**
Learning to use system	0,10	0,10	0,09	1,01	0,32
System capabilities	**0,38**	0,07	0,43	5,18	**<0,001**
Age	-0,01	0,02	-0,02	-0,27	0,79
Gender (0: male 1: female)	0,04	0,10	0,03	0,35	0,73
Computer literacy	0,01	0,06	0,01	0,17	0,87
Culture (0: German 1:Spanish students)	0,39	0,28	0,24	1,41	0,16
Major in college (0: Hum/soc 1:IT eng)	-0,34	0,31	-0,19	-1,12	0,27

R = 0,75 R^2 = 0,56 (adjusted R^2 = 0,52)

Increasing positive evaluation of the terminology and the system capabilities increase the overall reaction to the software. None of the background variables modify this relationship, however, even though gender and major in college does have an independent effect on system capabilities as shown in table 5.

3. CONCLUSION

The empirical testing presented in this paper shows that the background characteristics of the participants (including cultural background – native speakers of German vs. Spanish) do not predict the overall reaction to the software used in the simulation / role playing game (summarized in figure 3).

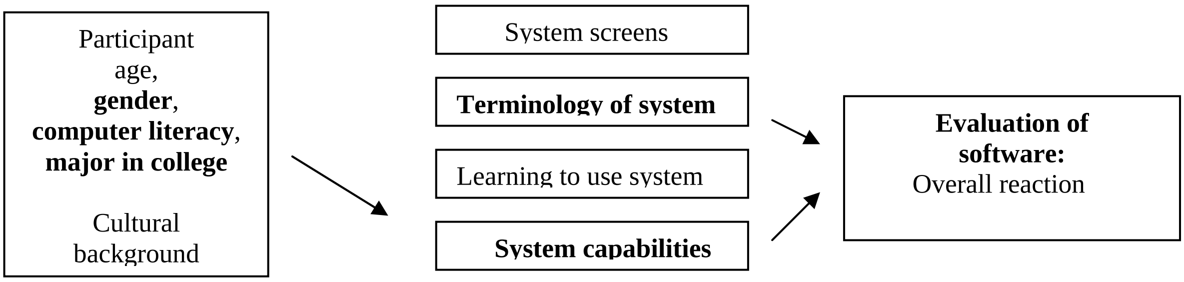

Figure 3. Modified research model: evaluation of software, cultural differences and participant characteristics

Of the four intervening variables (the evaluation of system components), the evaluation of the terminology and the evaluation of the system capabilities do have an effect on the overall reaction to the software.

This research has shown that the evaluation of the software used in the simulation / role playing game is not affected by the cultural background of the participants, nor by age. Gender, major in college (humanities and social sciences vs. IT/engineering students) and computer literacy do have an effect on just one of the intermediary variables – the evaluation of the system capabilities. IT/engineering students have more experience with computer systems and thereby expect a more streamlined software system. None of the background variables have an effect on the three other intermediary variables.

In order to improve the software used for the IDEELS simulations it will be necessary to focus on upgrading the system terminology (e.g. error messages and prompts for input) and system capabilities (e.g. system speed and reliability).

ACKNOWLEDGEMENT

Project Intercultural Dimensions in European Education through on-Line Simulation (IDEELS - http://www.ideels.uni-bremen.de/) was funded as a CDI and CDID project by the Socrates/Erasmus program of the European Commission.

REFERENCES

Ajzen, I. and Fishbein, M. 1980. *Understanding Attitudes and Predicting Social Behaviour.* Prentice-Hall, Englewood Cliffs, NJ, USA.

Chin, J. P. et al, 1988. Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. In *Proceedings of ACM CHI'88 Conference on Human Factors in Computing Systems*, pp. 213-218

Davis, F.D. 1993. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. In *International Journal of Man-Machine Studies*, No. 38. pp. 475-487.

Dix, Alan et al, 2004. *Human-Computer Interaction*. Pearson Education Limited, Essex, England.

Houle, P. A. 1996. Toward understanding student differences in a computer skills course. In *Journal of Educational Computing Research*. Vol 14. No. 1, pp. 25-48.

Joshi, K. 1992. A causal path model of the overall user attitudes toward the MIS function. The case of user information satisfaction. In *Information and Management*, No. 22, pp. 77-88.

Mahmood, Mo Adam, 2003. *Advanced Topics in End User Computing*. Idea Group Publishing, Hershey, PA, USA.

Orr, Claudia et al, 2003. *The Effect of Individual Differences on Computer Attitudes*. In Mahmood, Mo Adam, 2003, pp. 210-232.

Prince, Brad, 2004. Measures and Relationships of Computer Usage, Computer Locus of Control, Computer Literacy, and End-User Satisfaction among College of Business Seniors. *Proceedings of the 7th Annual Conference of the Southern Association for Information Systems.* Savannah, Georgia, USA, pp. 272-277.

Shneiderman, Ben, 1987. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Publishing Company, Reading, Mass., USA.

Sutherland, Janet, et al, 2003. Fostering Community in Networked Environments: IDEELS Telematic Simulations. *MENU conference 8-9 May, 2003*. Valencia, Spain. URL: http://www.upv.es/menuconf/