2005

Assessing the Reliability, Validity and Adaptability of PSSUQ

Ann Fruhling
University of Nebraska at Omaha, afruhling@mail.unomaha.edu

Sang Lee
University of Nebraska at Omaha, slee1@mail.unl.edu

Assessing the Reliability, Validity and Adaptability of PSSUQ

Ann Fruhling
University of Nebraska at Omaha
afruhling@mail.unomaha.edu

Sang Lee
University of Nebraska at Lincoln
slee1@mail.unl.edu

ABSTRACT

Psychometrically qualified usability evaluation instruments offer many advantages to the usability practitioner. Advantages include objectivity, replicability, quantification, economy, communication, and scientific generalization. It is important that instruments used in usability evaluation have shown acceptable estimates of reliability, validity, adaptability and practicality. This paper compares the psychometric properties of the Post-Study System Usability Questionnaire (PSSUQ) instrument in Lewis' study to a recent study that tailored the PSSUQ instrument to measure the user satisfaction of the usability of a web-based health provider interface.

Lewis reported that the PSSUQ had acceptable psychometric properties. However, Lewis' stated PSSUQ had limited generalizability and needed further examination. This research validated the PSSUQ instrument using a larger sample size in a different domain. The factor analysis and the clustering of the sub-scale items were different than Lewis' results. Nevertheless, this research concluded that the PSSUQ instrument is adaptable and produced solid psychometric results.

Keywords

Usability Evaluation, Instrument Validation, e-Health, Usability Testing

INTRODUCTION

Usability evaluation determines whether a system meets a pre-determined, quantifiable level of usability for specific types of users carrying out specific tasks. The self-administered questionnaire composed of multiple separate items organized into scales, with each scale assumed to measure an attribute or attitude dimension is a common approach to usability evaluation. Use of multiple items to assess each dimension is essential to the measurement process. When developing questionnaires it is necessary to examine the reliability and validity of the measurement instrument itself. Instrument validation studies are important because they 1) analyze the psychometric properties (reliability and validity) of an instrument and consequently the degree of confidence that can be placed in assertions based on that instrument and 2) document the results.

The PSSUQ instrument is intended to measure the system usefulness, information quality and interface quality of a user interface. The objective of this paper is to further validate the PSSUQ usability evaluation instrument and assess its adaptability to measure the user satisfaction of the usability of system interfaces in other domains such as telemedicine. The paper begins with an introduction of usability and instrument validation and also an overview of the PSSUQ instrument. The paper continues with a discussion of the psychometric details of a modified Post-Study System Usability Questionnaire (PSSUQ) referred to as Post-Study e-Health Usability Questionnaire (PSHUQ) which was used to evaluate the user satisfaction of the usability of a telemedicine system for a large mid-west medical center. Next, the psychometric results of the two instruments (PSSUQ and PSHUQ) are compared. The paper concludes with the implications from the results of the research.

BACKGROUND

Usability

Usability is a key to making systems easy to learn and easy to use (Nielsen and Mack, 1994). Usability includes the consistency and ease with which the user can manipulate and navigate the web site, clarity of interaction, ease of reading, arrangement of information, speed, and layout. Usability improves the design of user interfaces by evaluating the organization, presentation, and interactivity of the interface (Shneiderman, 1986). Prior research overwhelmingly suggests that usability is associated with many positive outcomes, such as a reduction in the number of errors, enhanced accuracy, a

more positive attitude on the part of the user toward the target system, and increased usage of the system by the user (Lecerof and Paterno 1998; Nielsen, 1996).

In order to develop computer systems with acceptable usability, it is vital for developers and designers to understand the factors that determine how people operate and make use of computer technology effectively and to translate that understanding into a system that will provide the capabilities for people to achieve efficient, effective, and safe interaction with the system interface. Usability is a key and proximal measurement for evaluating the success of an organization's web presence (Agarwal and Venkatesh, 2002).

Validity and Reliability of an Instrument

Straub (1989) states that instrument validation is a primary process in empirical research. Careful validation of the survey instrument can reduce measurement errors and increase validity. Instrument validation is concerned with content validity, construct validity, reliability, and criterion-related validity (Cooper and Schindler, 2001).

The characteristics of a sound measurement instrument are validity, which refers to the extent to which a test measures what one actually wishes to measure, reliability, which has to do with the accuracy and precision of the measurement procedure and finally, practicality, which addresses things such as economy, convenience and interpretability (Cooper and Schindler, 2001). An important point regarding validity and reliability is that a set of measurements can be reliable without being valid, but they cannot be valid without being reliable (Kachigan, 1991). Results are reliable if they are reproducible. If measurements on a set of items cannot be replicated, then the instrument is determined to be very unstable.

Content validity addresses the issue of whether instrument measures drawn are from all possible measures of the properties under investigation (Straub, 1989). In other words, do the items measure the content they were intended to measure (Creswell, 1994)? Content validity ensures that the operationalization of a construct adequately represents the domain of coverage of the construct. The typical procedures for assessing content validity are experts' assessments and subjects' assessments.

Construct validity refers to the correspondence between the results obtained from a measuring instrument and the meaning attributed to those results (Schwab, 1980). In other words, do the data reflect true scores or are they artifacts of the kind of instrument chosen (Straub, 1989)? Or in other words, do items measure hypothetical constructs or concepts (Creswell, 1994)? Factor analysis is one of the most common means to test construct validity (Schwab, 1980).

Factor analysis is a statistical procedure that examines the correlations among variables to discover clusters of related variables (Kachigan, 1991). It is common to do factor analysis because it provides summarized information which is sometimes easier to interpret. Since factor analysis identifies clusters of variables that are highly correlated with each other, it is possible to choose one variable from the set of potential predictor variables and thereby reducing collinearity problems (Kachigan, 1991).

Reliability measures the stability of the methodologies and is often thought of as the degree to which the measure supplies consistent results. In other words, are the item responses consistent across constructs (Creswell, 1994)? The internal consistency and the reliability of the scales are confirmed using inter-item correlations and Cronbach's coefficient alpha (Cronbach, 1951). Cronbach's coefficient alpha estimates the internal consistency of a set of items and indicates the extent to which the scale items belong to a common core construct.

PSSUQ

The Post-Study System Usability Questionnaire (PSSUQ) is a research instrument that was developed for use in scenario-based usability evaluation at IBM (Lewis, 1995). A follow-up study on the PSSUQ using five years of data produced similar psychometric properties between the original survey and the follow-up study survey (Lewis, 2002). The environment that the original PSSUQ instrument used at IBM to assess the user satisfaction of the usability was enterprise-wide and networked office application suites. The follow-up study domain was speech recognition systems.

PSSUQ consists of 19 items aimed to address the following five system usability characteristics: quick completion of work, ease of learning, high-quality documentation and online information, functional adequacy and rapid acquisition of usability experts and several different user groups (Lewis, 2002) that were identified by a panel of IBM HCI experts. See Table 1.

Post-Study System Usability Questionnaire Items
1. Overall, I am satisfied with how easy it is to use this system.
2. It was simple to use this system.
3. I could effectively complete the tasks and scenarios using this system.
4. I was able to complete the tasks and scenarios quickly using this system.
5. I was able to efficiently complete the tasks and scenarios using this system.
6. I felt comfortable using this system.
7. It was easy to learn to use this system.
8. I believe I could become productive quickly using this system.
9. The system gave error messages that clearly told me how to fix problems.
10. Whenever I made a mistake using the system, I could recover easily and quickly.
11. The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear.
12. It was easy to find the information I needed.
13. The information provided for the system was easy to understand.
14. The information was effective in helping me complete the tasks and scenarios.
15. The organization of information on the system screens was clear.
16. The interface of this system was pleasant.
17. I liked using the interface of this system.
18. This system has all the functions and capabilities I expect it to have.
19. Overall, I am satisfied with this system.

Table 1. Post-Study System Usability Questionnaire Items

Factor analysis of these 19 items from Lewis' research indicated there are three factor sub-scales. Lewis named the sub-scales: System Usefulness, Information Quality, and Interface Quality. The System Usefulness sub-scale refers to a system that is easy to use and easy to learn, allows the user to effectively complete tasks, and allows the user to become productive quickly. The Information Quality refers to the feedback the system provides to the user such as error messages and information on how to fix problems. It also includes information such as online help, onscreen messages and documentation that is clearly presented. Moreover, it also measures if the information is easy to understand, effectively helps the user complete tasks, and is organized. The Interface Quality sub-scale deals with how pleasant the system was to the consumer. It measures if s/he liked the system and if the system has all the functionality and capabilities s/he expected.

It was reported that the three sub-scales accounted for 87% of the variability in the data. The reliability using coefficient alpha analyses calculated that the reliability the sub-scales ranged from .91 to .96 as shown in Table 2. The correlation analyses supported the validity of the scales. Also, the sensitivity ANOVAs for the three sub-scales indicated significant differences among the groups. Therefore, from the results of the factor and reliability analyses it is reasonable for the PSSUQ to define three sub-scales (Lewis, 1995).

	Reliability
Factor 1 – System Usefulness	.96
Factor 2 – Information Quality	.91
Factor 3 - Interface Quality	.91

Table 2. PSSUQ Reliability Statistics

RESEARCH METHODOLOGY

PSHUQ

The objective of this research was to evaluate the validity, reliability and adaptability of the PSSUQ instrument in another setting. The PSSUQ instrument questions were modified to evaluate the usability of an e-health interface. The survey questions of the modified instrument called Post-Study e-Health Usability Questionnaire (PSHUQ) were reviewed by a panel of HCI experts for content validity and also evaluated on good survey question procedures, (i.e. only asking on question, asking a question that is concrete). Questions were changed to reference the e-health system the consumers were evaluating.

The rationale for changing "interface" to "e-health system" was that the subjects in this study are not familiar with the term "interface" and would be confused. This change provided more clarity. Table 3 shows a mapping of the PSHUQ questions to the original PSSUQ questions. Note that PSSUQ questions 3, 4, and 5 were combined in one PSHUQ question. The reason for this was their similarity and to reduce the number of survey questions. Unknown at the time of this research, Lewis was also examining the similarity of questions 3, 4, and 5 and statistically found that only using the results of question 4 had minimal effect on the coefficient alpha (Lewis, 2002). Therefore, this decision aligns with Lewis' latest thinking. Also, PSSUQ question 11 was split into two PSHUQ questions (9 and 10). The reasoning for this was that the original PSSUQ question was a compound question. In other words, it did not focus on one specific item, which is poor survey design.

PSHUQ	PSSUQ
1. Overall, I am satisfied with how easy it is to use this health web site.	1. Overall, I am satisfied with how easy it is to use this system.
2. This health web site was simple to use.	2. It was simple to use this system.
3a. I easily completed the task asked of me using this health web site.	3. I could effectively complete the tasks and scenarios using this system.
3b. I easily completed the task asked of me using this health web site.	4. I was able to complete the tasks and scenarios quickly using this system.
3c. I easily completed the task asked of me using this health web site.	5. I was able to efficiently complete the tasks and scenarios using this system.
4. I felt comfortable using this health web site.	6. I felt comfortable using this system.
5. This health web site was easy to learn how to use.	7. It was easy to learn to use this system.
6. I could become productive quickly using this health web site.	8. I believe I could become productive quickly using this system.
7. This health web site gave error messages that clearly told me how to fix problems.	9. The system gave error messages that clearly told me how to fix problems.
8. Whenever I made a mistake using the health web site, I could recover quickly.	10. Whenever I made a mistake using the system, I could recover easily and quickly.
9. The on-screen messages provided were clear.	11a. The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear.
10. The online help information provided was clear.	11b. The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear.
11a. It was easy to find the information I needed.	12. It was easy to find the information I needed.
11b. It was easy to find the information I needed.	13. The organization of information on the system screens was clear.
12. The information provided by the e-health web	14. The information provided for the system

site was easy to understand.	was easy to understand.
13. The information provided by the health web site helped me complete the tasks.	15. The information was effective in helping me complete the tasks and scenarios.
14. This health web site was pleasant.	16. The interface of this system was pleasant.
15. I liked using this health web site.	17. I liked using the interface of this system.
16. This health web site has all the capabilities I expected it to have.	18. This system has all the functions and capabilities I expect it to have.
17. Overall, I am satisfied with this health web site.	19. Overall, I am satisfied with this system.

Table 3. Mapping of Two Sets of Questions

Pilot Study

Following recommended practice, a pilot test was done that focused on clarity of instructions to participate in the study, content, readability, and understandability of the questionnaire, and the process of using the telemedicine web-site. A total of 60 students voluntarily participated in the study.

The factor analysis results in the pilot study indicated only two sub-scales: System Usefulness and Information Quality. This was different that the results of the PSSUQ factor analysis discussed earlier. It was speculated that the differentiation between System Interface Quality and System Information Quality was not distinguishable and therefore the two sub-scales collapsed into one. One possibility why this happened was the lack of errors and error messages the participants encountered. It was noted that the feedback of the participants and also after reviewing the data indicated that few error situations occurred. The two PSHUQ questions addressing error handling of the interface were often either left blank or comments were made that no errors occurred. Items dealing with errors are sub-items of the Information Quality factor in PSSUQ instrument. This feedback from the pilot study was taken into account and a N/A option was added so that blank answers would not result in the field study and to reduce possible frustration of the participant. Introducing potential errors for the participant to encounter was not an option. This e-health web-site was live and introducing errors would damage the reputation of the institution.

Field Study

The participants in the field study were rural consumers who are potential users of e-health care services. The data collection process consisted of a direct-mail invitation letter to rural residents asking them to participate in an e-health study via the Internet. The respondent was asked to browse around the e-health web site to become familiar with its contents and capabilities and then s/he was directed to various links for him or her to participate in the study. Also, participants were asked to complete a health risk assessment and then complete the online usability survey linked to the e-health web-site.

Response Rate

Out of the 7,850 invitation letters sent, 276 respondents voluntarily participated. Of those 276 participants, 266 responses provided data that were complete and matched the generalizing population. At first glance this may seem like a very low response, but one must take into account that only 42% of the rural population in this study has access to the Internet (IANR News Service, 2002). Therefore, the number of households that could indeed participate was 3,674. The response rate calculated to 7.13%.

Data Preparation

The data were analyzed for missing responses, incorrectly typed information such as the zip code, and making sure the resident did in fact live in a rural area. Upon further examination of incomplete data it was determined that items 13, 14, and 16 were only answered by 150 of the 266 respondents. This was not a surprise from the feedback of the pilot test because these items were error-related questions.

Another objective of data preparation is to test for normality of the scores of each of the variables. A normal distribution is characterized by *skewness*, a measure of the distribution's deviation from symmetry, and *kurtosis*, a measure of a distribution's peakedness or flatness (Cooper and Schindler, 2001). Scores with absolute values greater than 3.0 are described as being extremely non-normal. Absolute values of kurtosis greater than 10.0 may suggest a problem and scores greater than 20.0 may indicate an even more serious problem. After reviewing the frequency of responses statistics and skew graphs of the variables it was determined that the data passed the Normality Test.

Factor Analysis

Two factors with eigenvalues greater than one were extracted which accounted for 71.81 percent of the total variance. The first statistical Factor explains 64.75 percent of the variability, and the addition of the second Factor explains an additional 7.06 percent of the variability shown in Table 5.

	Eigenvalues	**% of Variance**
Factor 1 – System Quality	11.008	64.754
Factor 2 – System Usefulness	1.201	7.065

Table 4. Factor Analysis Results of System Usability Scale

The factor analysis from the field study produced similar results as the pilot study. Factor One included both of the PSSUQ interface quality and information quality subscales and was renamed to "System Quality." Factor Two corresponded with the PSSUQ System Usefulness sub-scale and kept the same name. These findings are different than the three factors Lewis (1995) identified. The compression of the interface quality and information quality sub-scales into one sub-scale could be related to the fact that both subscales evaluate quality, but from two different perspectives (e.g., information and interface) and yet, these two perspectives are not distinctive. This suggests that the instrument may not be sensitive enough. Lewis notes that the stage of the development the product is in and the type of product are thought to affect the sensitivity of the factor analysis (Lewis, 2002). This research supports that notion. Other reasons for this result may be the incomplete responses on the error-related survey questions and the high quality of the e-health web-site studied.

All Cronbach's coefficient alphas exceeded the generally accepted minimum value of .70, demonstrating satisfactory evidence of internal consistency. In fact, the two factors in this study were greater than .90 (see Table 5). The System Quality Reliability Coefficient Alphas is greater than both the PSSUQ Information Quality (.91) and the PSSUQ Interface Quality (.91) Reliability Coefficient Alphas. The System Usefulness Reliability Coefficient Alpha (.958) was very close to the PSSUQ System Usefulness Reliability Coefficient Alpha (.96).

	Reliability
Factor – System Quality	.933
Factor – System Usefulness	.958

Table 5. Summary of Cronbach's Coefficient Alpha Reliability from Factor Analysis

The sub-scale factor loading values are shown in Table 6. The factor loading value was .5. When two factors loaded on both factors the hirer factor loading value was selected. Table 6 also presents the corrected item-total correlation each sub-scale.

Sub-scale items	SysUsefulness Factor Loading	Quality Factor Loading	Item-Total Correlation
1. Overall, I am satisfied with how easy it is to use this health web site.	**.721**	.372	.794
2. This health web site was simple to use.	**.802**	.327	.776
3. I easily completed the task asked of me using this health web site.	**.894**	.186	.849
4. I felt comfortable using this health web site.	**.754**	.356	.798
5. This health web site was easy to learn how to use.	**.877**	.293	.896
6. I could become productive quickly using this health web site.	.374	**.786**	.844
7. This health web site gave error messages that clearly told me how to fix problems.	-.003	**.535**	.437
8. Whenever I made a mistake using the health web site, I could recover quickly.	.462	**.637**	.777
9. The on-screen messages provided were clear.	.571	**.643**	.838
10. The online help information provided was clear.	.572	**.586**	.789
11. It was easy to find the information I needed.	.555	**.618**	.804
12. The information provided by the health web site helped me complete the tasks.	.462	**.772**	.877
13. The information provided by the e-health web site was easy to understand.	.518	**.746**	.887
14. This health web site was pleasant.	.604	**.666**	.857
15. I liked using this health web site.	.470	**.672**	.787
16. This health web site has all the capabilities I expected it to have.	.332	**.744**	.784
17. Overall, I am satisfied with this health web site.	.401	**.844**	.903

Table 6. Factor Loadings and Item-Total Correlations

Further analysis shows that one subscale item in this study clustered differently than in Lewis' (1995) study. Table 7 shows how the items were grouped in Lewis' (1995) study compared to how the groupings emerged in this research. Question 6, "I could become productive quickly using this health web site" factored into the Quality sub-factor instead of System Usefulness. The e-health web-site in this study has won several awards on its presentation quality and perhaps the presentation quality of the system improved the user's productivity and influenced his or her response. And, therefore, the subject's responses mapped with system quality sub-scales.

Sub-scale items	This Study	Lewis (1995)
1. Overall, I am satisfied with how easy it is to use this health web site.	SysUsefulness	SysUsefulness
2. This health web site was simple to use.	SysUsefulness	SysUsefulness
3. I easily completed the task asked of me using this health web site.	SysUsefulness	SysUsefulness
4. I felt comfortable using this health web site.	SysUsefulness	SysUsefulness
5. This health web site was easy to learn how to use.	SysUsefulness	SysUsefulness
6. I could become productive quickly using this health web site.	**Quality**	SysUsefulness
7. This health web site gave error messages that clearly told me how to fix problems.	Quality	InfoQual

8. Whenever I made a mistake using the health web site, I could recover quickly.	Quality	InfoQual
9. The on-screen messages provided were clear.	Quality	InfoQual
10. The online help information provided was clear.	Quality	InfoQual
11. It was easy to find the information I needed.	Quality	InfoQual
12. The information provided by the health web site helped me complete the tasks.	Quality	InfoQual
13. The information provided by the e-health web site was easy to understand.	Quality	InfoQual
14. This health web site was pleasant.	Quality	InterQual
15. I liked using this health web site.	Quality	InterQual
16. This health web site has all the capabilities I expected it to have.	Quality	InterQual
17. Overall, I am satisfied with this health web site.	Quality	InterQual

Table 7. Sub-Item Clustering Comparison

Table 8 provides a qualitative description of the sub-scale items in terms of their means, standard deviation, and variance. Each construct was measured using a 7 point scale starting with a minimum value of zero, with "3" being neutral.

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
System Quality	266	0	6	4.2948	.06441	1.188
System Usefulness	266	0	6	4.7353	.06683	1.104

Table 8. Descriptive Statistics

Skewness is the measure of the symmetry of a distribution of data compared to a normal distribution. Substantial skewness is indicated by values falling outside the range of -1 and +1 (Groebner, Shannon, Fry, Smith, Groebner, 1993). The results in Table 9 indicate a slight presence of skewness in the data. Kurtosis is a measure of the "peakedness" or "flatness" of the distribution. The absolute value of 3.0 which is a statistical marker when the curve becomes non-normal and there is some cause for concern. System Usefulness exceeds this statistical marker, however System Quality and the Kurtosis of the entire dataset are acceptable.

	N Statistic	Skewness	Skewness	Kurtosis	Kurtosis
		Statistic	Std. Error	Statistic	Std. Error
System Quality	266	-.639	.149	1.036	.298
System Usefulness	266	-1.379	.149	3.153	.298
Overall Total	266	-.808	.149	1.407	.298

Table 9. Descriptive Statistics – Skewness and Kurtosis

CONCLUSION

A significant contribution of this study to the human computer interaction usability research field is that it has further validated an instrument that can be used in usability evaluation studies. The research methodology in this study carefully followed standardized procedures on the data collection and analysis, (e.g. pilot study, data cleansing, data analysis, etc.). The findings in this study show that the high reliability of the PSSUQ instrument was replicated both in a pilot study and in a field study using a larger sample and a different system application environment. Although, there was a difference in the number of sub-scale factors, this did not reduce the reliability and internal consistency.

Regarding content validity, the items in the PSSUQ instrument cover the breadth of attributes that are commonly associated with good usability practices. The five system characteristics that the questionnaire addressed were determined by IBM

usability experts and human factors experts. Content validity for the PSHUQ was examined by a panel of experts and survey questions were adjusted to good survey design. These changes did not negatively affect the reliability of the PSHUQ instrument.

In addition, this paper addresses the adaptability of a usability evaluation instrument that initially evaluated the user satisfaction of the usability of office application suites. In this study the instrument was adapted to evaluate the user satisfaction of the usability of an e-health website. The psychometric results of the PSSUQ instrument were comparable to the PSHUQ instrument with the exception of the factor analysis. Two possible explanations for this difference is 1.) the primitiveness of the user interfaces that Lewis used in the initial study compared to today's sophisticated interfaces and 2.) the lack of errors and error messages encountered by the subjects in the study. Therefore, the instrument may not to be sensitive enough and thus, produced different factor analysis results in the new domain.

A further contribution of this study is the introduction of the PSHUQ instrument. E-health care practitioners can use this instrument to assess the usability of an e-health web site. This study serves as a benchmark for future e-health care practitioners and researchers.

A limitation of this research is that the PSHUQ survey was cross-sectional, that is the survey information was collected at one point in time. Another limitation of this study, although a positive aspect of the system, was the lack of errors that occurred in the system to the user. This reduced the completeness of the survey responses and may have also affected the sensitivity of the instrument. In conclusion, it is reasonable to state that PSSUQ is a reliable, adaptive instrument that can be used to evaluate the user satisfaction of the usability of various systems.

REFERENCES

1. Agarwal, R. and Venkatesh, V. (2002) Assessing a firm's web presence: a heuristic evaluation procedure for the measurement of usability, *Information Systems Research,* June, 13, 2, 168-186.
2. Cooper, D. R. and Schindler, P. S. (2001) *Business Research Methods,* McGraw-Hill Irwin.
3. Creswell, J. (1994) *Research Design, Qualitative and Quantitative Approaches,* SAGE Publications.
4. Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests," *Psychometrika*, September, 16, 297-334.
5. Groebner, D., Shannon, P., Fry, P., Smith, K. (2000) *Business Statistics, A Decision-making Approach*, Prentice Hall.
6. IANR News Service (2002) More Rural Nebraskans Using Computers, Internet Technologies NU Poll Shows, Institute of Agriculture and Natural Resources, UNL, October 22.
7. Kachigan, S. K.(1991) *Multivariate Statistical Analysis: A Conceptual Introduction,* Radius Press, New York.
8. Lacerof, A. and Paterno, F. (1998) Automatic support for usability evaluation, *IEEE Transaction Software Engineering*¸24, 863-887.
9. Lewis, J. R. (1995) IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use, *International Journal of Human-Computer Interaction,* 7, 1, 57-78.
10. Lewis, J. R. (2002) Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies, *International Journal of Human-Computer Interaction,* 14(3&4), 463-488.
11. Nielsen, J. and Mack, R. (1994) *Usability inspection methods,* John Wiley & Sons, Inc.
12. Nielsen, J. (1996) Usability metrics: Tracking interface improvements, *IEEE Software,* November, 12-13.
13. Schwab, D.P. (1980) Construct validity in organization behavior, In B.M. & L.L. Cummings (Eds.), *Research in organizational behavior*, Vol. 2. Greenwich, CT: JAI Press.
14. Shneiderman, B. (1986, 1987c.) *Designing the User Interface: Strategies for Effective Human-Computer Interaction,* Reading, Mass., Addison-Wesley.
15. Straub, D. W. (1989) Validating instruments in MIS research, *MIS Quarterly,* June, 147–169.