Development of a Customizable Health IT Usability Evaluation Scale

Po-Yin Yen, RN, PhD[1], Dean Wantland, RN, PhD[1], Suzanne Bakken, RN, DNSc[1,2]
[1]School of Nursing & [2]Dept of Biomedical Informatics, Columbia University, NY, NY

Abstract

We developed a customizable questionnaire, the Health Information Technology (IT) Usability Evaluation Scale (Health-ITUES) and conducted an exploratory factor analysis to examine the scale's psychometric properties. Nurses (n=377) completed Health-ITUES to rate the usability of a web-based communication system for scheduling nursing staff. The analysis revealed a four-factor structure of Health-ITUES. The results provided preliminary evidence for the factorial validity and internal consistency reliability of Health-ITUES.

Keywords

Usability evaluation, scale development

Introduction

Psychometric properties have been reported for a number of instruments designed to measure user perceptions of system usability including: IBM Computer System Usability Questionnaire [1], Technology Acceptance Model (TAM) Perceived Usefulness/Ease of Use , Unified Theory of Acceptance and Use of Technology (UTAUT) [2], Questionnaire for User Interaction Satisfaction (QUIS) [3], Physician Order Entry User Satisfaction and Usage Survey [4], End-User Computing Satisfaction [5].

Although validated instruments exist, it was reported that a mismatch between study needs and concepts measured in the questionnaires resulting in item addition, deletion, or modification[6]. In addition, most often questionnaires fail to consider "task" as a variable in the questionnaire even though "task" has been demonstrated to be essential for health IT usability evaluation[7-8].

To address these knowledge gaps, we developed the Health-ITUES considering tasks by addressing various levels of expectation. In this paper we describe scale development and report the initial psychometric assessment of Health-ITUES.

Background

We designed the scale items within the context of evaluating a specific system, a web-based communication system for scheduling nursing staff. Consequently, a description of the system precedes the description of scale development.

The web-based communication system

The web-based communication system, Bidshift, allows nurse managers to announce open shifts throughout their organization and staff nurses to request shifts for which they are qualified based upon their profile. If more than one nurse requests the same open shift, nurse managers are able to select a nurse based on her/his experience or working hours (not exceeding hospital overtime policy) for patient safety purposes. The primary goal of the web-based communication system is to improve the efficiency and effectiveness of the staffing and scheduling process.

Prior to developing Health-ITUES, we conducted two preliminary usability studies related to the web-based communication system. The first study was conducted in Bryn Mawr Hospital, a component of Main Line Hospital System, a site in which the web-based communication system had been implemented for two years. User-system interaction was studied using a think-aloud protocol[9]. The study found that patterns of system use varied among individuals during usability testing. Their usage of each function depended on how effective they perceived the function to be in achieving a particular task. Users were satisfied when a function was both effective and efficient for accomplishing a task. We also conducted heuristic evaluation[10] with human-computer interaction experts to examine usability issues from the perspective of Nielsen's usability heuristics[10] and identified minor interface design issues. These preliminary studies enhanced our familiarity with the web-based communication system and informed the development of system-specific outcome items for Health-TUES.

Scale Development

Health-ITUES development was iterative. First, Main Line Hospital and the web-based communication system developer, Concerro, Inc., proposed items based upon existing questionnaires. Second, we conceptually mapped the proposed items to the subjective measures of Health IT Usability Evaluation Model (Health-ITUEM), which is an integrated model that we developed based on multiple theories to include both subjective and objective measures for usability evaluation[11]. We identified which Health-ITUEM constructs and concepts were missing in proposed items. Third, considering technology acceptance as the subjective measure of usability evaluation, we added items from TAM measurements of perceived usefulness and perceived ease of use[12] and IBM Computer System Usability Questionnaire[1] to represent missing constructs and concepts.

Based upon the principle that usability is developed through user-centered design methods but measured through the interaction of user, tool, and task in a specified setting[8, 13-14], we modified items to address the web-based communication system and specific user tasks. We identified the task-specific concepts in each question for modification. For example, to modify an original TAM question, "Using [system] is useful in my job", we identified the system by name and also specified user tasks. The resulting question was "Using Bidshift (system) is useful for requesting shifts (task)".

Also of note, in contrast to most satisfaction measures that report general information which cannot identify specific usability problems[15], Health-ITUES items address different levels of expectation. These include: a). task level: "I am satisfied with Bidshift for requesting open shifts", b). individual level: "The addition of BidShift has improved my job satisfaction", and c). organizational level: "BidShift technology is an important part of our staffing process".

The final Health-ITUES consists of 36 items rated on a 5-point Likert scale from strongly disagree to strongly agree: "actual usage" (2 items), "intention to use" (1 item), "satisfaction" (5 items), "perceived usefulness" (6 items), "perceived ease of use" (3 items), "perceived performance speed" (2 items), "learnability" (2 items), "competency" (2 items), "flexibility/customizability" (3 items), "memorability" (2 items), "error prevention" (2 items), "information needs" (3 items), and "other outcomes" (3 items). Higher scale value indicates higher perceived usability of the technology.

Research Questions

- What is the factorial structure of Health-ITUES?
- What is the internal consistency reliability of Health-ITUES factors?

Methods

A cross-sectional study design was used to evaluate users' perception toward the web-based communication system after the system implementation.

Setting and Sample

Main Line Health comprises six hospitals in the Philadelphia area. The study was conducted at Bryn Mawr Hospital, which has approximately 1500 staff nurses. The web-based communication system had been implemented for two years at the time of the evaluation. All staff nurses were qualified to participate in the study.

Sample size

A minimum sample of 5 to 10 observations per scale item is recommended for factor analysis[16]. Kaiser-Meyer-Olkin test (KMO-test) and Bartlett's test of sphericity were also performed to examine the adequacy of sample size for exploratory factor analysis.

Data collection procedure

Surveys were electronically distributed to the hospital staff via email. An announcement was also posted on the login page for the web-based communication system. The period of data collection was four weeks.

Data analysis

We included only Question (Q)3 to Q35 for analysis because Q1, Q2 and Q36 are questions assessing self-report intention to use and actual usage, which were used for evaluating the predictive validity of Health-ITUES.

SPSS 16.0 and Mplus 5.21 were used for data analysis. Mplus is an application which provides features and functions for factor analysis and structural equation modeling. Exploratory factor analysis was used to explore the psychometric characteristics of the scale. We first examined item communalities. Next, we performed parallel analysis (PA) and Velicer's Minimum Average Partial (MAP) test to determine the number of factors extracted. PA tends to underestimate the number of components while Velicer's MAP test tends to overestimate[17-19]. Therefore, the optimal decision is likely to be made after both methods were performed[18].

Analytic procedures included: 1) PA and Velicer's MAP to determine the number of components; 2) Maximum Likelihood (ML) robust extraction method (also called Satorra-Bentler method), that is recommended for non-normal distributed data[20]; 3) orthogonal (varimax) and oblique rotations (promax) to assess stability of the factor solution across rotation types; and 4) item reductions based upon item loadings (.32 or higher on two or more factors or less than half the difference of factor loading with other factors)[16, 21] and affect on Cronbach's alpha reliabilities. Following item reduction, we repeated procedures until the final solution was reached.

Results

Descriptive Analysis

Health-ITUES respondents included 377 staff nurses. The majority of respondents rated themselves as Internet competent. All entries were completed without missing data. The sample of 377 met the rule-of-thumb for minimum sample size for factor analysis[16]. Sample size adequacy was also supported by the KMO-test value of .964. Bartlett's test of sphericity supported the appropriateness of the data for factor analysis (Approx. χ^2=11363.162, df=528, p=0.000).

Selection of Number of Factors for Extraction

The results of the PA suggested extraction of three factors. In contrast, the results of Velicer's MAP test

and an examination of factors with eigenvalues greater than one supported the extraction of four factors. Consequently, we decided to extract four factors.

Exploratory Factor Analysis

The average communality greater than 0.6 indicated strong influence by an underlying construct. We conducted the exploratory factor analysis using robust ML. Promax and varimax rotations showed similar loading solutions with the exception of Q 8, which loaded in Factor 4 in the promax rotation, but in Factor 2 in the varimax rotation. We conducted further analysis based upon the promax solution.

Factor Naming

"Quality of work life" comprises 6 items (Q34, Q33, Q35, Q32, Q18, Q19) that characterize system impact beyond the system functionality. For example, Q33 and Q35 relate to organization and staffing processes.

"Perceived Usefulness" comprises 12 items (Q29, Q26, Q28, Q30, Q25, Q31, Q21, Q27, Q14, Q20, Q15, Q24) that assess system usefulness for a targeted task, requesting shift. For example, Q29, Q28, Q26, Q30 and Q27 are all from TAM's perceived usefulness items. Other items include those related to efficiency (Q20 and Q21), information needs (Q14 and Q15), system satisfaction (Q31) and ease of use item (Q24).

"Perceived Ease of Use" comprises 8 items (Q5, Q4, Q6, Q22, Q23, Q10, Q11, Q3) focused on evaluating user-system interaction. For example, Q5 and Q6 are indicators of competency; Q22 and Q23 are indicators of ease of use.

"User Control" comprises 7 items Q12, Q13, Q16, Q7, Q8, Q9, Q17) related to user control ability. For example, Q12 and Q13 are indicators of error prevention function; Q16 asks about information needs to minimize difficulty performing the system. Customizability (Q7, Q8, Q9), assesses system capacity for adjusting to users' various operation preference or habits.

Item Deletion

We deleted items based on factor loadings less than .32 on all factors (e.g., Q24), item-loadings at .32 or higher on two or more factors (e.g. Q15, Q20, Q23, Q11), or cross-loadings less than half the difference from an item's highest factor loading (e.g., Q3, Q8, Q9, Q18, Q19). Deletion of Q17 improved the Cronbach's alpha reliability. These 11 deletions resulted in a 22-item Health-ITUES.

We performed a second exploratory factor analysis using robust ML with promax rotation to assess the stability of the factor structure. Question 7 and Q32 cross-loaded on Factors 2 and 3 and Factors 1 and 2, respectively, and were subsequently removed. There were no cross-loadings in a third exploratory factor analysis. The final factor structure is displayed in Table 1. Cronbach's alphas ranged from .81 to .95. Factor correlations ranged from 0.37 to 0.66. The final Health-ITUES in generic form, i.e., system, task, and user not specified is shown in Table 2.

Table 1. Factor Loadings

	QWL	PU	PEU	UC
Q34	0.831			
Q33	0.749			
Q35	0.618			
Q29		0.921		
Q26		0.865		
Q28		0.815		
Q25		0.733		
Q27		0.695		
Q30		0.691		
Q31		0.673		
Q21		0.649		
Q14		0.552		
Q5			0.959	
Q4			0.938	
Q6			0.910	
Q22			0.717	
Q10			0.482	
Q12				0.907
Q13				0.813
Q16				0.492

Note: QWL=Quality of Work Life; PU=Perceived Usefulness; PEU=Perceived Ease of Use; UC=User Control

Discussion

The final Health-ITUES included 20 items. The analyses supported the factorial validity of Health-ITUES. Internal consistency reliability was excellent. Several items were deleted because of cross-loading on two or more factors thus suggesting that the items may require modification to be more explicitly tied to a factor. For example, Q3 asks "I received sufficient education to log on to and use Bidshift" and cross-loaded on "Perceived ease of use" and "User Control". Q3 originally was intended to measure learnability, which influences system ease of use. This may reflect the correlation between education and user control.

Customizability items Q7, Q8 and Q9, which were taken from TAM and IBM measures [1, 12], were deleted because of cross-loading issues. Possible reasons include lack of specificity in the items and the fact that the customizability of the web-based communication system was not clear which may have resulted in mixed user opinions.

Table 2. Health IT Usability Evaluation Scale

Item		Concept
Quality of Work Life (Cronbach α= .94)		
34	I think [BidShift] has been [a positive addition to Nursing].	System impact-career mission
33	I think BidShift has been [a positive addition to our organization].	System impact-organizational level
35	[BidShift technology] is an important part of [our staffing process].	System impact-personal level
Perceived Usefulness (Cronbach α= .94)		
29	Using [Bidshift] makes it easier to [request the shift I want].	Productiveness
26	Using [Bidshift] enables me to [request shifts] more quickly.	Productiveness
28	Using [Bidshift] makes it more likely that I [will be awarded a shift that I request].	Productiveness
30	Using [Bidshift] is useful for [requesting open shifts].	General usefulness
25	I think [Bidshift] presents a more equitable process for [requesting open shifts].	General usefulness
31	I am satisfied with [Bidshift] for [requesting open shifts].	General satisfaction
21	I [am awarded shifts] in a timely manner because of [Bidshift].	Performance speed
27	Using [Bidshift] increases [requesting open shifts].	Productiveness
14	I am able to [find shifts that I am qualified to work] whenever I use [Bidshift].	Information needs
Perceived Ease of Use (Cronbach α= .95)		
5	I am comfortable with my ability to use [Bidshift].	Competency
4	Learning to operate [Bidshift] is easy for me.	Learnability
6	It is easy for me to become skillful at using [Bidshift].	Competency
22	I find [Bidshift] easy to use.	Ease of use
10	I can always remember how to log on to and use [Bidshift].	Memorability
User Control (Cronbach α= .81)		
12	[Bidshift] gives error messages that clearly tell me how to fix problems.	Error Prevention
13	Whenever I make a mistake using [Bidshift], I recover easily and quickly.	Error Prevention
16	The information (such as on-line help, on-screen messages and other documentation) provided with [Bidshift] is clear.	Information needs

Items related to "Other outcomes" (Q17, Q18, Q19) (i.e. Q17: "Since BidShift, I receive fewer calls at home asking me to come in") have lower means (3.41, 3.60 and 3.59 respectively) compared to the mean score (4.06) of all items. These items were proposed by the system developer and organization in which the system was implemented and represented their visions for system impact. The lower mean scores suggest their expectations of system impact may not match those of system users under real world conditions.

Existing questionnaires tended to support user-system interaction assessment (e.g. the system is easy to use), evaluate general satisfaction (e.g. the system is useful to my job), or target a specific system (e.g. the system helps me to be efficient at medication administration) rather than address various levels of expectation. If health IT supports achievement of specific tasks, but does not impact higher level expectations such as job satisfaction, user acceptance may be variable.

In this study, exploratory factor analysis clearly defined factors associated with levels of expectation. "User Control" and "Perceived Ease of Use" capture user-system interaction, while "Perceived Usefulness" evaluates task accomplishment through system use, and "Quality of Work Life" represents higher expectations of system impact. The study demonstrated that level of expectation reflect both simple tasks and higher system impact. The customizability allows adaptation to various health IT characteristics and has the potential to provide comparison across similar systems in the future.

Limitations

The study has some limitations. First, the response rate for the survey was only about 25%. While the sample size was deemed sufficient for factor analysis by the KMO-test value, the results of the factor analysis might vary with broader population representation. Second, most participants were competent in Internet use. The results may vary in users with low internet competency. However, the user variance should be minimized with user training.

The high internet competency could be seen as the intended outcome after user training. Finally, the study was conducted using only one system and one professional group, registered nurses which may potentially limit broader applicability. However, Health-ITUES was designed to be customizable based on the user-system-task-environment interaction. Future work will apply Health-ITUES to other health IT or other professions to address this potential limitation.

Conclusion

We developed a customizable questionnaire (Health-ITUES) for measuring perceived Health IT usability. The results of exploratory factor analysis provided preliminary evidence for the factorial validity and internal consistency reliability of Health-ITUES.

Acknowledgement

References

1. Lewis JR. IBM Computer Usability Satisfaction Questionnaires - Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction.* 1995;7(1):57-78.
2. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. *Mis Quarterly.* 2003;27(3):425-78.
3. Norman K, Shneiderman B. Questionnaire for User Interaction Satisfaction (QUIS). 1997; http://lap.umd.edu/quis/, 2008.
4. Lee F, Teich JM, Spurr CD, Bates DW. Implementation of physician order entry: user satisfaction and self-reported usage patterns. *Journal of the American Medical Informatics Association.* 1996;3(1):42-55.
5. Doll WJ, Xia WD, Torkzadeh G. A Confirmatory Factor-Analysis of the End-User Computing Satisfaction Instrument. *Mis Quarterly.* 1994;18(4):453-461.
6. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. *Journal of Biomedical Informatics.* 2010;43(1):159-172.
7. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. *BMC Medical Informatics & Decision Making.* 2006;6:3.
8. Shackel B. Usability—context, framework, definition, design and evaluation. In: Shackel B, Richardson SJ, eds. *Human factors for informatics usability.* New York, NY: Cambridge University Press; 1991:21-37.
9. Yen P, Bakken S. Usability Testing of a Web-based Tool for Managing Open Shifts on Nursing Units. Paper presented at: The 10th International Congress on Nursing Informatics 2009; Helsinki, Finland.
10. Yen P, Bakken S. A Comparison of Usability Evaluation Methods: Heuristic Evaluation versus End-User Think-Aloud Protocol - An Example from a Web-based Communication Tool for Nurse Scheduling. *AMIA Annu Symp Proc.* 2009;Annual Symposium Proceedings/AMIA Symposium.
11. Schnall R, Hyun S, Yen P, Bakken S. Using Technology Acceptance Models to Inform the Design and Evaluation of Nursing Informatics Innovations (Panels). *Proceedings of NI2009, the 10th International Congress in Nursing Informatics.* June 28 - July 1, 2009.
12. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly.* 1989;13(3):318-340.
13. Bennett J. *Visual Display Terminals: Usability Issues and Health Concerns.* Englewood Cliffs New Jersey: Prentice-Hall; 1984.
14. ISO 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability. 1998.
15. Hornbaek K. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies.* Feb 2006;64(2):79-102.
16. Worthington RL, Whittaker TA. Scale development research - A content analysis and recommendations for best practices. *Counseling Psychologist.* 2006;34(6):806-838.
17. Zwick WR, Velicer WF. Comparison of 5 Rules for Determining the Number of Components to Retain. *Psychological Bulletin.* 1986;99(3):432-442.
18. O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods Instruments & Computers.* 2000;32(3):396-402.
19. Turner NE. The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement.* 1998;58(4):541-568.
20. Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika.* 2001;66(4):507-514.
21. Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Reseach & Evaluation.* 2005;10(7):1-8.