PSYCHOMETRIC EVALUATION OF THE POST-STUDY SYSTEM USABILITY QUESTIONNAIRE: THE PSSUQ

James R. Lewis
IBM Design Center / Human Factors
Boca Raton, FL

Usability evaluators used an 18-item, post-study questionnaire in three related usability tests. I conducted an exploratory factor analysis to investigate statistical justification to combine items into subscales. The factor analysis indicated that three factors accounted for 87 percent of the total variance. Coefficient alpha analyses showed that the reliability of the overall summative scale was .97, and ranged from .91 to .96 for the three subscales. In the sensitivity analyses, the overall scale and all three subscales detected significant differences among the user groups; and one subscale indicated a significant system effect. Correlation analyses support the validity of the scales. The overall scale correlated highly with the sum of the After-Scenario Questionnaire ratings that participants gave after each scenario. The overall scale also correlated moderately with the percentage of successful scenario completion. These results are consistent with the hypothesis that these alternative measurements tap into a common underlying construct. This construct is probably usability, based on the content of the questionnaire items and the measurement context.

INTRODUCTION

The purpose of this paper is to describe a psychometric evaluation of a questionnaire (the Post-Study System Usability Questionnaire, or PSSUQ) that assesses user satisfaction after participation in scenario-based usability studies. Psychometric instruments that evaluate computer-user satisfaction are not new (Ives, Olson, and Baroudi, 1983; LaLomia and Sidowski, 1990). However, none of the previous scales were specifically developed for usability studies. The questionnaire items are constituent items for summative (Likert) scales (McIver and Carmines, 1981; Nunnally, 1983). This paper will cover item construction, item selection, exploratory factor analysis of the items, estimates of scale reliability, assessment of scale sensitivity, and estimates of construct validity. (It is outside the scope of this paper to address alternatives to summative scaling, such as the magnitude scaling approach of Cordes, 1984.)

ITEM CHARACTERISTICS

The items are seven-point graphic scales, anchored at the end points with the terms "Strongly agree" for 1 and "Strongly disagree" for 7, and a "Not Applicable" (N/A) point outside the scale. A group of usability evaluators selected the items on the basis of their comprehensive content regarding hypothesized constituents of usability. For example, the items assess such system characteristics as ease of use, ease of learning, simplicity, effectiveness, information, and the user interface. (See Appendix A for the questionnaire items.)

PSYCHOMETRIC EVALUATION

The Usability Studies

Lewis, Henry, and Mack (1990) conducted studies to evaluate the usability characteristics of three office application systems. Forty-eight employees of temporary help agencies participated in the studies. The three office systems included a word processor, a mail application, a calendar application, and a spreadsheet on three different platforms (computer hardware and operating systems) that allowed a certain amount of integration among the applications. We assessed the systems with eight common office benchmark scenarios (Lewis, Henry and Mack, 1990). We used the 3-item After-Scenario Questionnaire (Lewis, 1991) and the 18-item Post-Study System Usability Questionnaire (Lewis, 1990) to measure participant attitude. All of the questionnaire items were seven-point scales. We used a between-subjects design, with each participant assigned to one system. After a 30-minute system-exploration period, participants performed the scenarios (randomly ordered) with their assigned system, and completed the After-Scenario Questionnaire (ASQ) (Lewis, 1991a; Lewis, 1991b) after each scenario. After participants completed all scenarios, they rated the system using the Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 1990).

Results

The scree plot (see Cliff, 1987, p. 313) for the principal factors analysis indicated that a three-factor solution was appropriate. Table 1 shows the varimax-rotated factor pattern. Bold type in Table 1 highlights factor loadings that exceed 0.5. Two items (2 and 12) loaded highly on two factors, so I did not include them in any potential subscales.

Table 1. Varimax-rotated factor pattern for the principal
factor pattern of the PSSUQ

Item	Scale	Factor 1	Factor 2	Factor 3
1	3	0.22	0.28	**0.80**
2	N/A	**0.58**	0.22	**0.64**
3	1	**0.77**	0.26	0.43
4	1	**0.63**	0.35	0.46
5	1	**0.75**	0.38	0.25
6	1	**0.81**	0.45	0.07
7	1	**0.80**	0.16	0.36
8	1	**0.68**	0.38	0.48
9	1	**0.69**	0.46	0.40
10	3	0.30	0.36	**0.75**
11	3	0.37	0.36	**0.76**
12	N/A	0.30	**0.59**	**0.56**
13	2	0.05	**0.61**	0.24
14	2	0.36	**0.71**	0.24
15	2	0.45	**0.63**	0.25
16	2	0.44	**0.75**	0.22
17	2	0.43	**0.70**	0.32
18	2	0.43	**0.74**	0.40

Based on the item content, a group of usability
evaluators named the scales 1: System Usefulness
(SYSUSE), 2: Information Quality (INFOQUAL) and 3:
Interface Quality (INTERQUAL). The factor analysis
indicated that the three factors accounted for 87 percent of
the total variance. Coefficient alpha analyses showed that
the reliability of the overall summative scale was .97, and
ranged from .91 to .96 for the three subscales (SYSUSE=.96,
INFOQUAL=.91, and INTERQUAL=.91). In the sensitivity
analyses, the overall scale and all three subscales detected
significant differences among the user groups (OVERALL:
$F(2,29)=4.35$, $p=.02$; SYSUSE: $F(2,36)=6.9$, $p=.003$;
INFOQUAL: $F(2,33)=3.68$, $p=.04$; INTERQUAL:
$F(2,35)=3.74$, $p=.03$); and INFOQUAL showed a significant
system effect ($F(2,33)=3.18$, $p=.05$). Correlation analyses
support the validity of the scales. The overall scale correlated
highly with the sum of the After-Scenario Questionnaire
ratings that participants gave after each scenario ($r(20)=.80$,
$p=.0001$). The overall scale also correlated moderately with
the percentage of successful scenario completion ($r(29)=-
0.40$, $p=.026$). The SYSUSE ($r(36)=-0.40$, $p=.006$) and
INTERQUAL ($r(35)=-0.29$, $p=.08$) correlated moderately with
the percentage of successful scenario completion.

DISCUSSION

These results are consistent with the hypothesis that these
alternative measurements tap into a common underlying
construct. This construct is probably usability, based on the
content of the questionnaire items and the measurement
context. Others who conduct usability studies should
consider using this post-study questionnaire (with additional
items, if necessary).

REFERENCES

Cliff, N. Analyzing multivariate data. San Diego, CA:
Harcourt, Brace, Jovanovich.

Cordes, R. E. (1984). Use of magnitude estimation for
evaluating product ease-of-use (In E. Grandjean (Ed.)
International Scientific Conference on Ergonomic and
Health Aspects of Modern Offices). London: Taylor
and Francis.

Ives, B., Olson, M., and Baroudi, J. (1983). The
measurement of user satisfaction. Communications
of the ACM, 26, 785-793.

LaLomia, M. J. and Sidowski, J. B. (1990). Measurements of
computer anxiety, attitudes, literacy and aptitudes: A
review. International Journal of Human-Computer
Interaction, 3, 231-253.

Lewis, J. R. (1990). Psychometric evaluation of a post-study
system usability questionnaire: The PSSUQ (Tech.
Report 54.535). Boca Raton, FL: International
Business Machines Corp.

Lewis, J. R. (1991a). An after-scenario questionnaire for
usability studies: Psychometric evaluation over three
trials. SIGCHI Bulletin, 23, 79.

Lewis, J. R. (1991b). Psychometric evaluation of an after-
scenario questionnaire for computer usability studies:
The ASQ. SIGCHI Bulletin, 23, 78-81.

Lewis, J. R., Henry, S. C., and Mack, R. L. (1990).
Integrated office software benchmarks: A case
study. In Human-Computer Interaction -- INTERACT
'90 (pp. 337-343). London: North-Holland.

McIver, J. P. and Carmines, E. G. (1981). Unidimensional
scaling. Sage University Paper series on
Quantitative Applications in the Social Sciences,
series no. 07-024. Beverly Hills, CA: Sage
Publications.

Nunnally, J. C. (1978). Psychometric theory. New York, NY:
McGraw-Hill.

Appendix A. The Post-Study System Usability Questionnaire

This questionnaire is designed to give you an opportunity to tell us your reactions to the system you used. Your responses will help us understand what aspects of the system you are particularly concerned about, and the aspects with which you are satisfied. To as great a degree as possible, think about all the tasks that you have performed with the system while you answer these questions.

Instructions

Read each statement and indicate HOW STRONGLY YOU AGREE OR DISAGREE WITH THE STATEMENT by circling a number on the scale. If a statement does not apply to you, circle N/A. We encourage you to write comments to elaborate on your answers. After you complete this questionnaire, we'll go over your responses with you to make sure we understand all of your responses. Thank you!

1.   This system has all the functions and capabilities I expect it to have.

Strongly agree   Strongly disagree   Not Applicable

   1   2   3   4   5   6   7   N/A

Comments:

2.   Overall, I am satisfied with this system.

Strongly agree   Strongly disagree   Not Applicable

   1   2   3   4   5   6   7   N/A

Comments:

3.   Overall, I am satisfied with how easy it is to use this system.

Strongly agree   Strongly disagree   Not Applicable

   1   2   3   4   5   6   7   N/A

Comments:

4.   It was simple to use this system.

Strongly agree   Strongly disagree   Not Applicable

   1   2   3   4   5   6   7   N/A

Comments:

5.   I could effectively complete the tasks and scenarios using this system.

Strongly agree   Strongly disagree   Not Applicable

   1   2   3   4   5   6   7   N/A

Comments:

6.   I was able to complete the tasks and scenarios quickly using this system.

Strongly agree   Strongly disagree   Not Applicable

        1   2   3   4   5   6   7   N/A

        Comments:

7.   I was able to efficiently complete the tasks and scenarios using this system.

Strongly agree   Strongly disagree .   Not Applicable

        1   2   3   4   5   6   7   N/A

        Comments:

8.   I felt comfortable using this system.

Strongly agree   Strongly disagree   Not Applicable

        1   2   3   4   5   6   7   N/A

        Comments:

9.   It was easy to learn to use this system.

Strongly agree   Strongly disagree   Not Applicable

        1   2   3   4   5   6   7   N/A

        Comments:

Note: The interface includes those items that you use to interact with the system, such as the keyboard, mouse, screens, graphics and language.

10.   The interface of this system was pleasant..

Strongly agree   Strongly disagree   Not Applicable

        1   2   3   4   5   6   7   N/A

        Comments:

11.   I liked using the interface of this system.

Strongly agree   Strongly disagree   Not Applicable

        1   2   3   4   5   6   7   N/A

        Comments:

12.   The organization of information on the system screens was clear.

Strongly agree   Strongly disagree   Not Applicable

        1   2   3   4   5   6   7   N/A

        Comments:

13.   The system gave error messages that clearly told me how to fix problems.

Strongly agree   Strongly disagree   Not Applicable

    1   2   3   4   5   6   7   N/A

Comments:

14.   Whenever I made a mistake using the system, I could recover easily and quickly.

Strongly agree   Strongly disagree   Not Applicable

    1   2   3   4   5   6   7   N/A

Comments:

15.   The information provided with this system (on-line help, documentation) was clear.

Strongly agree   Strongly disagree   Not Applicable

    1   2   3   4   5   6   7   N/A

Comments:

16.   It was easy to find the information I needed.

Strongly agree   Strongly disagree   Not Applicable

    1   2   3   4   5   6   7   N/A

Comments:

17.   The information provided for the system was easy to understand.

Strongly agree   Strongly disagree   Not Applicable

    1   2   3   4   5   6   7   N/A

Comments:

18.   The information was effective in helping me complete the tasks and scenarios.

Strongly agree   Strongly disagree   Not Applicable

    1   2   3   4   5   6   7   N/A

Comments: