Research Note

On the Test-Retest Reliability of Perceived Usefulness and Perceived Ease of Use Scales

By: **Anthony R. Hendrickson**
Computer Information Systems and Quantitative Analysis
University of Arkansas
BADM 204
Fayetteville, Arkansas 72701 U.S.A.

Patti D. Massey
Computer Information Systems and Quantitative Analysis
University of Arkansas
BADM 204
Fayetteville, Arkansas 72701 U.S.A.

Timothy Paul Cronan
Computer Information Systems and Quantitative Analysis
University of Arkansas
BADM 204
Fayetteville, Arkansas 72701 U.S.A.

As information technology (IT) continues to provide organizational decision makers with a greater abundance and assortment of information systems (IS), the need for valid and reliable instruments to assess the success of these systems is increasingly important (Jarvenpaa, et al., 1985; Straub, 1989). An underlying tenet of IS success is the decision maker's willingness to adopt and utilize these systems. Measures that predict and explain use are important in determining what causes people to accept or reject information technology. This note reports on the test-retest reliability of the perceived usefulness and perceived ease of use scales.

Davis (1989) developed and validated two scales for assessing user acceptance of information technology—perceived usefulness and perceived ease of use. Adams, et al. (1992) replicated the work of Davis (1989) to demonstrate the validity and reliability of these scales. While replicating the study, Adams, et al. (1992) also extended it to different settings and found both scales to have the same validity and reliability characteristics as the Davis (1989) study. Using two different samples, they demonstrated the internal consistency and replication reliability of the two scales.

An instrument's reliability can be examined by using tests of internal consistency, replication with different samples, and test-retest using the same sample. Replication using different samples can be used to determine convergent and discriminant validity. The stability of an instrument can be determined using a test-retest procedure. The test-retest method involves multiple administrations of an instrument to the same people to assess the instrument's consistency and reliability. The key difference between test-retest and replication is the use of the same subject group for multiple instrument administrations. Theoretically, this eliminates any potential confounding due to heterogenous subjects.[1] The test-retest method of reliability has been applied to other measurement instruments used in MIS research (Galletta and Lederer, 1989; Hawk and Raju, 1989; Torkzadeh and Doll, 1991).

Although Adams, et al. (1992) reported on the results of the validity and reliability of the perceived usefulness and perceived ease of use

[1] Three potential problems must be addressed when applying the test-retest methodology: recall, time, and reactivity (Nunnally, 1978). A recall problem may arise when subjects are administered the instrument within too short an interval. Subjects may recall their responses and respond based on recall; recalled responses will affect the instrument's ability to produce consistent results. Similarly, a time problem may arise if the subjects are administered the instrument within too long an interval; differing subject responses may be attributed to changes in the subjects themselves and not inconsistencies in the instrument. Lastly, a problem with reactivity can occur when subjects are administered the instrument multiple times. Subjects become sensitized to the instrument and "learn" to respond as they perceive they are expected to respond (Nunnally, 1978).

scales, the test-retest reliability of these scales has not been reported. Our study examined the test-retest reliability of the perceived usefulness and perceived ease of use scales, using two software packages, adding further evidence concerning the reliability of these scales.

Perceived Usefulness and Perceived Ease of Use: Test-Retest Reliability

Perceived usefulness is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" (Davis, 1989, p. 320). Perceived ease of use refers to "the degree to which a person believes that using a particular system would be free of effort" (Davis, 1989, p. 320). These two definitions were used to generate the 12 items that define the constructs in the Davis (1989) perceived usefulness and perceived ease of use scales. Two studies (one using a file editor and electronic mail and the other using two graphics software packages) were conducted to assess the internal validity and reliability of the scales.

Subjects for the study were undergraduate students in a major midwestern university enrolled in an introductory computing course in which software instruction is part of the course curriculum. The instrument was administered to two samples—one using a spreadsheet package[2] with 51 subjects and one using a database management package[3] with 72 subjects. The information systems experience of the subjects varied; some had little or no previous information systems experience, while others had experience on mainframe and personal computer systems.

The perceived usefulness and perceived ease of use scales were administered to both samples using the same test-retest methodology. Subjects were given the instrument at two separate points in time. The instrument was administered (T1) after a short period of introductory instruction (functionality and application) and use of each software package. The second administration (T2) followed the first administration by three days to assess the reliability of the instrument. During the three-day interval, subjects continued to use the software with no additional instruction. Therefore, most variance was across subjects rather than applications.

To assess the consistency and reliability of the scales, three methods were used—Cronbach's alpha, paired t-test, and correlation. For each sample, Cronbach's alpha coefficients were calculated for the perceived usefulness and perceived ease of use subscales for both administrations of the instrument. Paired t-tests were conducted to determine the difference in mean responses. Finally, Spearman correlation coefficients were calculated to assess the reliability of the individual items and subscales for both samples.

Test-Retest Results

Davis (1989) reported alpha coefficients of .98 for perceived usefulness and .94 for perceived ease of use subscales. The alpha coefficients for both applications in this study are presented in Table 1. Alpha coefficients for both samples are comparable to the results obtained by Davis (1989). The smallest alpha coefficient found was .89, while the largest was .96. Based on the results of this study, the instrument is reliable. In light of the concise number of items employed in the Davis (1989) instrument, these alpha coefficients are extremely good indicators of the instrument's reliability.

The results of the paired t-tests of the subjects' mean responses and correlation coefficients between individual scale items for the sample using the spreadsheet package are shown in Table 2. Correlations between the individual scale items for this sample range from a low of .54 to a high of .73. There are two differences for the individual item mean scores (T1-T2) that are statistically different at the 0.05 level—"Easier to do Work" and "Useful in Work." No significant differences were found between the subscale means.

Similarly, the results of the paired t-tests of the subjects' mean responses and correlation coefficients between individual scale items for the

[2] Lotus 1-2-3 was selected due to its near universal acceptance and use in business and industry. In addition, Lotus 1-2-3 was one of the software packages used in the replication study by Adams, et al. (1992).

[3] Paradox 3.5 was also selected due to its use in business and industry.

Table 1. Test-Retest Reliability: Cronbach's Alpha

	Spreadsheet Application	Database Application
Initial Administration		
Usefulness	.89	.94
Ease of Use	.90	.93
Second Administration		
Usefulness	.95	.96
Ease of Use	.93	.94

sample using the database management package are shown in Table 3. Correlations between the individual scale items for this sample range from .58 to .79. In this sample, there are two individual items that have significantly different means (T1-T2)—"Enhance Effectiveness" and "Easy to Learn." In addition, there are no differences between the subscale means. For both samples, only one subscale correlation was less than .80. Since Nunnally (1978) suggests that correlations of .80 or higher are very good, the subscales are reliable.

Conclusion

While the perceived usefulness and perceived ease of use individual scale item correlation results are not exceptionally high, the subscale correlations are very high. This, combined with the minimal number of significant mean differences for items, indicates the test-retest reliability of the Davis (1989) instrument.

These results are consistent with previous test-retest results for instruments measuring the success of information systems (Galletta and Lederer, 1989; Torkzadeh and Doll, 1991). This is especially the case if the mean scores and alpha coefficients are considered. The risk of extending the reliability results of this study is reduced because two software packages were analyzed. However, some caution should be taken when generalizing these results to applications for which the instrument has not been validated. Nonetheless, the results found in this study demonstrate that the Davis (1989) instrument exhibits a high degree of test-retest reliability.

References

Adams, D.A., Nelson, R.R., and Todd, P.A. "Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication," *MIS Quarterly* (16:2), June 1992, pp. 227-247.

Davis, F. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:4), September 1989, pp. 319-340.

Galletta, D.F. and Lederer, A.L. "Some Cautions on the Measurement of User Information Satisfaction," *Decision Sciences* (20:3), Summer 1989, pp. 419-438.

Hawk, S.R. and Raju, N.S. "Test-Retest Reliability of User Information Satisfaction: A Comment on Galletta and Lederer's Paper," *Decision Sciences* (22:5), November/December 1991, pp. 1165-1170.

Jarvenpaa, S.L., Dickson, G.W., and DeSanctis, G. "Methodological Issues in Experimental IS Research: Experiences and Recommendations," *MIS Quarterly* (9:2), June 1985, pp. 141-156.

Nunnally, J.C. *Psychometric Theory,* McGraw-Hill, New York, NY, 1978.

Straub, D.W. "Validating Instruments in MIS Research," *MIS Quarterly* (13:2), June 1989, pp. 147-166.

Torkzadeh, G. and Doll, W.J. "Test-Retest Reliability of the End-User Computing Satisfaction Instrument," *Decision Sciences* (22:1), Winter 1991, pp. 26-37.

Note: Scale Reliability

Table 2. Test-Retest Statistics: Spreadsheet Application

	Test Mean T1	Retest Mean T2	Correlation (T1-T2)*	Significance of t-value
Usefulness				
Accomplish Task Quickly	1.706	1.687	.62	.8211
Improve Performance	1.824	1.706	.68	.1823
Increase Productivity	1.745	1.725	.62	.8211
Enhance Effectiveness	1.706	1.608	.66	.2796
Easier to do Work	1.843	1.686	.73	.0443**
Useful in Work	1.420	1.686	.64	.0035**
Subscale Total	**1.690**	**1.683**	**.85**	**.7094**
Ease of Use				
Easy to Learn	2.020	1.961	.73	.4967
Easy to Manipulate	2.176	2.078	.69	.3583
Clear/Understandable Interaction	2.216	2.059	.54	.1583
Flexible to Interact With	2.140	2.098	.60	.7425
Easy to Become Skillful	1.765	1.882	.59	.2039
Easy to Use	1.902	2.020	.65	.2609
Subscale Total	**2.037**	**2.016**	**.77**	**.8373**

* Initial and second administration, all significant at 0.01 level.
** Statistically significant at 0.05 level.

Table 3. Test-Retest Statistics: Database Application

	Test Mean T1	Retest Mean T2	Correlation (T1-T2)*	Significance of t-value
Usefulness				
Accomplish Task Quickly	3.347	3.361	.78	.8957
Improve Performance	3.339	3.416	.76	.7876
Increase Productivity	3.306	3.401	.58	.4171
Enhance Effectiveness	3.139	3.431	.62	.0233**
Easier to do Work	3.486	3.375	.58	.4901
Useful in Work	3.347	3.583	.58	.0909
Subscale Total	**3.336**	**3.433**	**.81**	**.2972**
Ease of Use				
Easy to Learn	2.681	2.875	.76	.0258**
Easy to Manipulate	3.099	3.111	.63	.9052
Clear/Understandable Interaction	3.014	3.028	.67	.8898
Flexible to Interact With	3.153	3.083	.59	.5325
Easy to Become Skillful	2.611	2.736	.79	.1064
Easy to Use	2.778	2.903	.77	.1291
Subscale Total	**2.885**	**2.956**	**.86**	**.1878**

* Initial and second administration, all significant at 0.01 level.
** Statistically significant at 0.05 level.