ASAC 1995
Windsor, Ontario

Wynne W. Chin
Barbara Marcolin
Faculty of Management
University of Calgary

THE HOLISTIC APPROACH TO CONSTRUCT VALIDATION IN IS RESEARCH: EXAMPLES OF THE INTERPLAY BETWEEN THEORY AND MEASUREMENT[1]

The paper presents two empirical examples of how the reliability and validity of a set of measures can change contingent on the theoretical model (or lack thereof) in which they are to be applied. In particular, measures developed atheoretically where the constructs are tested in separate components analysis are shown not to work well in the context of a theoretical model. Furthermore, measures that are demonstrated to work well in one theoretical model are shown not to generalize to another context. Finally, the multi-dimensionality of a set of measures can often be masked using first generation analytic techniques whereas second generation methods such as Partial Least Squares can provide additional information. It is suggested that researchers consider using a "one step" approach for theory and measurement development. At a minimum, this paper recommends that final validation of the psychometric properties of a set of measures for a construct should be made in the context of the theory in which they are to be used.

Introduction

One indicator of maturity in a discipline, as in the case of the Information Systems (IS) community, is the amount of attention directed toward the methodological issues of measurement. It is argued that valid and reliable measures of concepts or constructs are a prerequisite to model testing or theory building (Zmud & Boynton, 1991; Straub, 1989; Klenke, 1992). As such, IS researchers tend to follow a two step approach where the measures for relevant constructs are developed first prior to their use in a theoretical context (e.g., Davis, 1989; Segars & Grover, 1993). Recent advances in computerized programs for latent variable modeling (e.g., LISREL, EQS, PLS) allow the researcher to follow an alternative approach where both theoretical and measurement models can be jointly estimated. Consistent with current philosophical positions (Suppe, 1977), these second generation multivariate techniques allow the researcher to transcend the doctrine of operationalism where a concept is synonymous with the corresponding set of operations or the positivistic notion of independence between theory and data. Instead, theory and data (i.e., empirical measurements) are seen to interact. To establish the validity of a set of measures involves analyzing them simultaneously in the theoretical context in which they are to be applied.

Fornell (1989, p. 157) argues that this interplay between theory and measures must occur and that a separation between the two is not possible. Specifically, he suggests that:

- Observation if it is to be relevant, must be interpreted.
- That in terms of which interpretation is made is always theory.
- The theory not only serves as a basis of interpretation, but also determines what is to be counted as an observation, problem, method, solution, and so forth.

[1] The authors wish to acknowledge Stephen Hayne for establishing contact with the organization and assisting in the delivery and receipt of the surveys.

This perspective therefore suggests that measures established as valid (via separate factor analysis) will not necessary remain so when applied with other constructs in the context of a theoretical model. Furthermore, even if established within the context of a theory, measures for a particular construct may not generalize when used in other theoretical models (Fornell & Yi, 1992). The purpose of this paper is, therefore, to provide an empirical demonstration of these two points; we show how the reliability and validity of measures can change contingent on the theoretical context (or lack thereof) in which they are used. In particular, we show how a measurement model for a construct with four indicators can differ in three contexts. The first context, representative of the traditional approach, is to test the measures without being embedded in a theoretical model. Here, only the epistemic relationships between the manifest indicators and construct are examined. In the second context, our focal construct is then causally connected to a second construct. The third context is obtained by causally connecting our focal construct with yet a third construct. We show how the parameter estimates linking the measures to our focal construct (i.e., factor loadings) can change contingent on the other construct to which it is linked (i.e., the nomological network).

We use two focal constructs to highlight our claims. The first construct (Voluntariness) represents a set of measures that have gone through an initial validation procedure (Moore and Benbasat, 1991), but it has yet to be widely used. We use these measures primarily to demonstrate how the measurement model can change dependent on whether it is used within a nomological net or not. We do not pursue further as to the reason for the changes in the reliability and thus validity of these measures. To understand and discuss the results requires more substantive knowledge. One potential reason is that a construct's set of indicators may not be uniformly tapping into the same underlying factor. It is possible that some of the indicators are influenced by another factor in addition to the one it was intended to measure. Alternatively, and to be demonstrated here, some of the measures may actually be measuring another factor and yet be masked within the components analysis. Thus, our second focal construct is a contrived one to highlight this multidimensional issue. We use two measures from each of two factors: Perceived Usefulness and Perceived Ease of Use (Davis, 1989). The measures for these two constructs have gone through a substantial amount of testing and validation (Davis 1989; Davis et al. 1989; Davis 1993; Mathieson 1991; Adams, Nelson and Todd, 1992; Hendrickson, Massey and Cronan 1993). The constructs, themselves, have been shown to be reasonably unidimensional. Thus, the use of these measures allows us greater ease in interpretation of the analyses.

We limit our analyses within the context of a components based approach using Partial Least Squares (Wold, 1982; Chin & Gopal, 1993; Barclay, Higgins, & Thompson, forthcoming). The primary reason is that principal components analysis has been and continues to be the most widely used method among factor analytic techniques in psychology (see Lilly, Hoaglin, & Anderson-Kulman, 1989, cited in Velicer & Jackson, 1990). As an indicator of it's prevalent use among IS researchers, we examined all articles published in MIS Quarterly from 1989 to 1994 inclusive. Of the 145 published articles, 14 of them used factor analytic techniques. Seven of them were specified as principal componets (3 used Varimax rotation, 1 used oblique rotation, 3 unspecified). The other seven articles did not name the specific technique beyond the term factor analysis (6 used varimax rotation, 1 unspecified). Thus, at least 50 percent of the articles (with an upward potential of 100 percent) used component analysis.

Method

Sample

The data set was obtained from a single organization that had recently installed voice mail. A total of 60 questions relating to a recent introduction of voice mail were presented. Of the 575 questionnaires distributed, 241 usable responses were returned representing a 42 percent response rate. On average, the respondents had been using voice mail for less than a month, sent 2.14 messages per

day (standard deviation- s.d. = 2.07) and received 4.49 messages per day (s.d. = 4.87). Respondents were on average 39 years old (s.d. = 9.28) and had worked for the company an average of 11 years (s.d. = 6.9). Sixty percent of the respondents were male. The respondents came from various levels in the organization, 13 percent were managers, 12 percent were engineers and 38 percent were technicians, and the remaining 37 were clerical workers.

Measures

The measures used in this study (listed in the Appendix) come from research involved with IT adoption and use. The construct definitions which were used in the formation of these items are presented in Table 1. For simplicity and ease of comparison, only the first two items for Perceived Usefulness and Perceived Ease of Use are used to develop the contrived multi-dimensional construct (hereafter referred to as Construct?).

Table 1. Construct Definitions

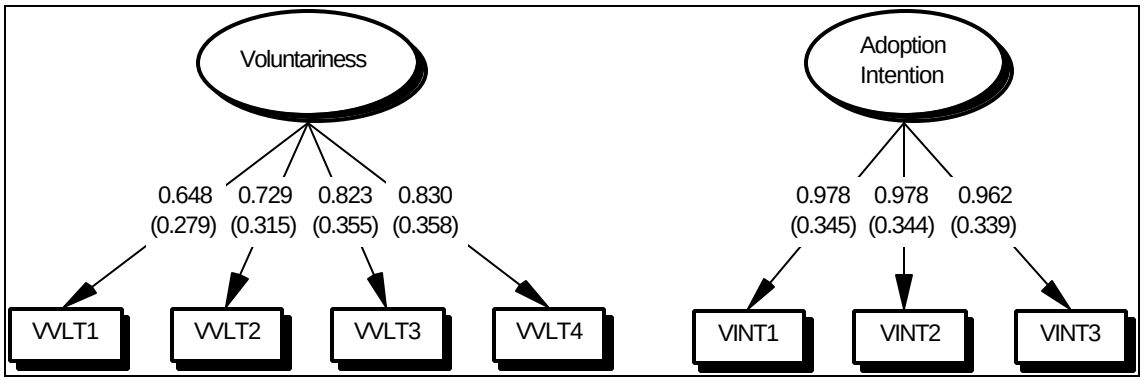
Constructs	Source	Original Definition
Perceived Usefulness	Davis (1989)	The degree to which a person believes that using a particular system would enhance his or her job performance.
Perceived Ease of Use	Davis (1989)	The degree to which a person believes that using a particular system would be free of effort.
Compatibility	Moore and Benbasat (1991)	The degree to which an innovation is perceived as being consistent with the existing values, needs, and past experiences of potential adopters.
Voluntariness	Moore and Benbasat (1991)	The degree to which use of the innovation is perceived as being voluntary, or of free will.
Result Demonstrability	Moore and Benbasat (1991)	The degree to which the results of an innovation are communicable to others.
Adoption intention	authors	A measure of the strength of one's intention to perform a behavior (e.g., use voice mail).

Empirical Example 1: Voluntariness

In our first analysis, we examine Moore and Benbasat's (1991) Voluntariness construct as it relates to Adoption Intention. Figure 1 provides the component loadings and weights obtained from our PLS analysis.[2] As depicted, the two constructs are kept separate without a causal connection. This analysis, thus, becomes equivalent to performing two separate principal components analysis. In the absence of a theoretical context linking Voluntariness to Adoption Intention, we would conclude that the measures for Adoption Intention are excellent (all above 0.96). In the case of our focal construct, Voluntariness, the third and fourth measures (VVLT3 and VVLT4) have respectible loadings above 0.8. The second indicator, VVLT2, is above the typical recommended cutoff of 0.7. The first indicator, with a loading of 0.648, would seem to be the least reliable of the measures. Thus, if we had to consider dropping one of these measures for reasons such as space constraints in a survey, we might opt for first indicator. The weights, as expected, parallel those of the loadings and therefore suggest the same conclusions.

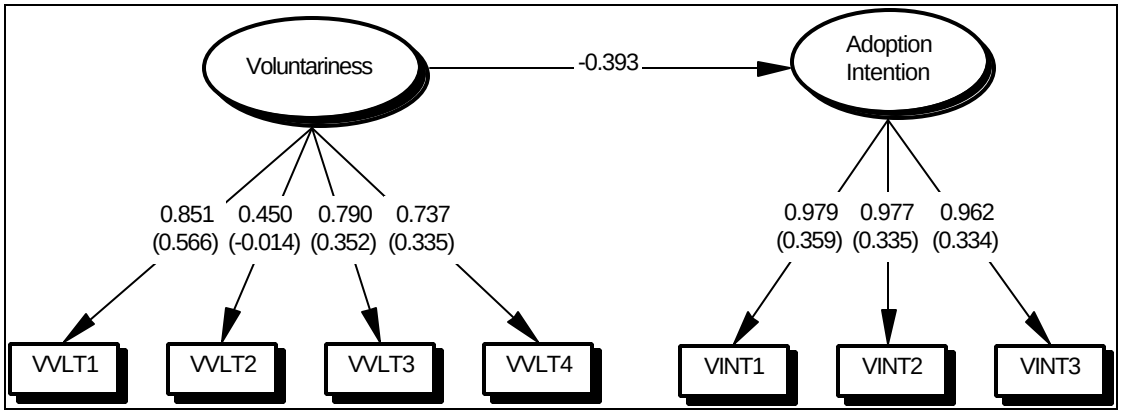
[2] Loadings are interpreted as the extent to which the measure is correlated with the overall component score. Weights refer to the degree to which a measure is used in creating an overall component scored.

Figure 1. PLS Loadings And Weights (In Parenthesis) Of Volutariness And Adoption Intention Where The Constructs Are Modeled As Theoretically Independent



The reliability of the loadings for Voluntariness change dramatically when we use the measures in a theoretical context. A second PLS analysis is conducted with the same measures, but with a theoretical linkage between Voluntariness and Adoption Intention. The results (Figure 2) show that in the context of predicting adoption intention, as opposed to in a separate components analysis, the first indicator suddenly become the most representative of Voluntariness with a loading of 0.851. The third and fourth indicators are above the 0.70 cutoff. But, surprisingly, the second indicator drops to a unreliable level of 0.45. Furthermore, we see that the weight for this measure drops to a negligible -0.014.

Figure 2. PLS Loadings And Weights (In Parenthesis) Of Volutariness And Adoption Intention Where The Constructs Are Modeled In A Causal Relationship



As a post hoc analysis, we might come up with a plausible rationale for why the first measure made such a dramatic improvement when used in the context of predicting adoption intention. We might conjecture that the 4 items are not necessarily measuring the same thing. The first measure seems more

indicative of normative expectations of one's superior. The second measure reflects both the influence of superiors as well as job description. The third and fourth measures use the verbs "require" and "compulsory" which may reflect a stronger and possibly different influence when compared to expectations used in the first measure. Thus, a person may be aware of a superior's expectation, yet not necessarily view it as a requirement. Whether this multidimensional explanation is true or not requires further research, but this is not the crux of this study. Rather, we have empirically demonstrated that the reliability (and thus utility for predictive purposes) of an indicator (i.e., VVLT1) can change when it is embedded in a theoretical context as opposed to a separate atheoretical components analysis.

Empirical Example 2: Contrived Factor

Building upon the results of our first empirical analysis, we continue our demonstration of the interplay between theory and measures by creating a contrived multi-dimensional factor. We use an equal number of measures from two different constructs to create a single factor. Using a traditional components analysis procedure, we demonstrate that a researcher may incorrectly consider the measures all tap into a single factor. But, when used in two different theoretical contexts, the reliability of these measures vary depending on the linkage to the second construct.

Assuming a researcher mistakenly used multidimensional measures to represent a single factor, the traditional approach to test for their reliability and convergent validity would be an exploratory components analysis. Using our contrived construct consisting of two indicators from Perceived Usefulness (VUSF1 and VUSF2) and two indicators from Perceived Ease of Use (VEOU1 and VEOU2), we performed a principal components analysis and two related methods (Harman, 1976, pp. 133-173). Table 2 presents the results where a single, albeit incorrect, factor was specified. Overall, the results may lead a researcher to conclude that the measures are reasonable with moderate reliability and seemingly reflect a single factor.

Table 2. Results Of From Three Factor Analytic Procedures

	Principal Components	Principal factors (centroid)	Principal factors (communality)
VUSF1	0.724	0.599	0.656
VUSF2	0.756	0.664	0.687
VEOU1	0.714	0.576	0.665
VEOU	0.77	0.70	0.728

We next perform a PLS analysis on both our contrived construct and the measures for Compatibility (Moore and Benbasat, 1991). Figure 3 presents the results. The loadings for Construct? is, as expected, the same as the results of the principal components analysis. Those for Compatibility are all quite high with weights approximately equal.

Figure 3. PLS Loadings And Weights (In Parenthesis) Of Compatibility And Contrived Construct? Where The Constructs Are Modeled As Theoretically Independent

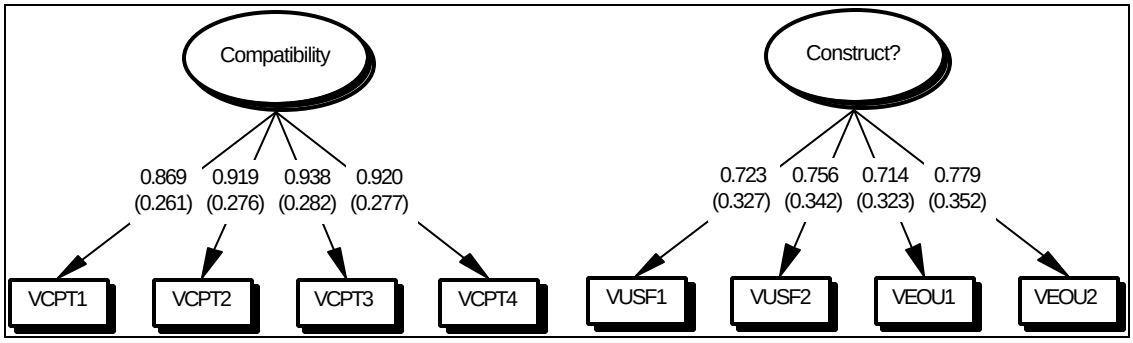
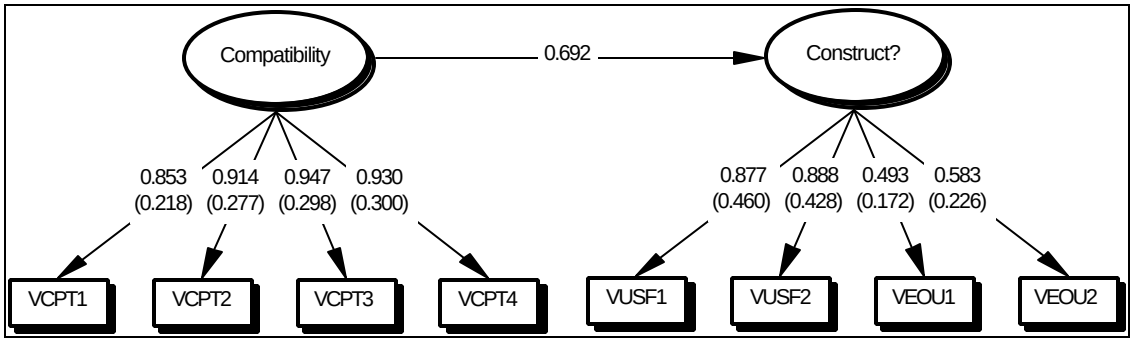


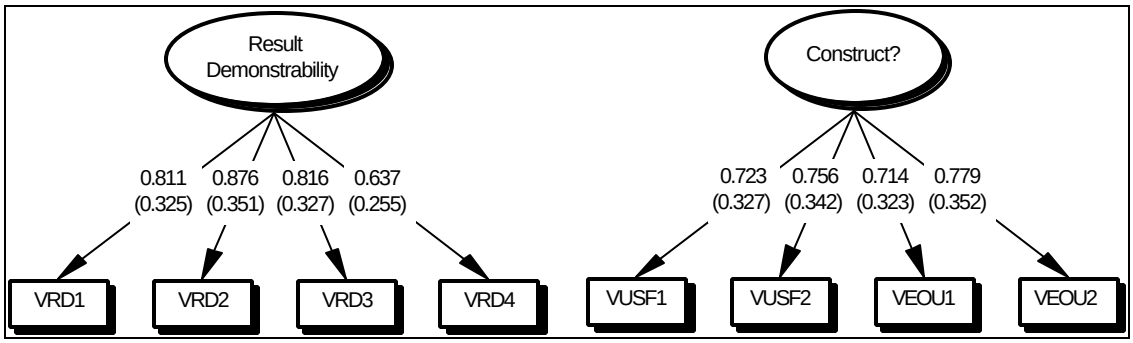
Figure 4 presents the results when Construct? is causally specified to be impacted by Compatibility. While the loadings and weights for Compatibility remains approximately the same, those for Construct? have changed dramatically. Overall, the differences in the loadings and weights among the indicators suggest that Construct? primarily represents Perceived Usefulness. In particular, we see that the items VUSF1 and VUSF2 have much higher loadings compared to those of VEOU1 and VEOU2. This is also reflected in the weights where over two thirds of the construct score is contributed by the Perceived Usefulness items. The results, therefore, seem to suggest that the items for Construct? are: 1) not unidimensional and 2) ideally represented by Perceived Usefulness when theoretically linked with Compatibility. Thus, we might conjecture that individuals when answering the questions about whether Voice Mail is compatible with their work are referring more to job performance gains from Voice Mail than its ease of use.

Figure 4. PLS loadings and weights (in parenthesis) of Compatibility and contrived Construct? where the constructs are modeled in a causal relationship.



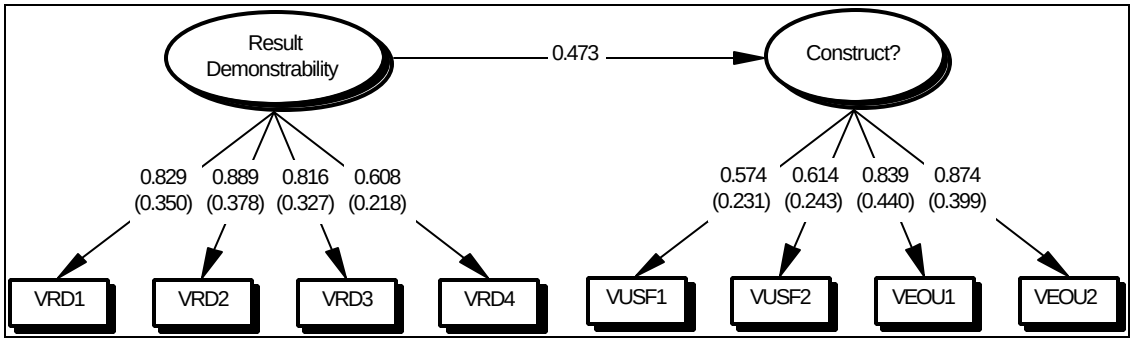
As a final example of how measures can change in the theoretical context of other constructs, we replace the Compatibility construct with Result Demonstrability. The results of the PLS analysis where the constructs are modeled independently is shown in Figure 5. The first three indicators for Result Demonstrability are quite good with loadings above 0.80. The fourth indicator is less reliable with a loading of 0.637. In the case of Construct?, the results are as before.

Figure 5. PLS Loadings And Weights (In Parenthesis) Of Result Demonstrability And Contrived Construct? Where The Constructs Are Modeled As Theoretically Independent



When the constructs are causally connected, the meaning of Construct? changes once more as reflected with the differentiation among the measures. In this case, as depicted in Figure 6, Construct? becomes primarily Perceived Ease of Use. As opposed to the earlier results using Compatibility, the loadings and weights for the Perceived Usefulness items now go down while those for Perceived Ease of Use are higher. Result Demonstrability, on the other hand, remains relatively the same. From these results, we might suggest that individuals answering the questions concerning whether that can communicate the results of using Voice Mail are primarily referring to the ease of use rather than performance gains.

Figure 6. PLS Loadings And Weights (In Parenthesis) Of Result Demonstrability And Contrived Construct? Where The Constructs Are Modeled In A Causal Relationship



Discussion and Conclusion

The results of our two empirical examples demonstrated how the reliability and validity of measures can change in the context of a theoretical model. In particular, measures developed atheoretically where the constructs are tested in separate components analysis do not necessarily imply that they will work well in the context of a theoretical model. Furthermore, measures that are shown to work well in one theoretical model will not necessarily translate equally to another context.

In demonstrating the first point, we showed how the reliability of measures for Voluntariness changed when used to predict Adoption Intention. We suggested that one reason may be the multidimensional nature of the measures. In a separate components analysis, only the epistemic relationship between the indicators and constructs are examined. But when a causal connection is made

between constructs, the appropriateness of a set of measures relate not only to how well they tap into a construct, but also how well they predict (or are predicted by) another construct.

In the second example, we once again show how the reliability and validity of measure tested without being embedded in a theoretical model can be deceiving. Using indicators from two different factors, we provided a demonstration of how traditional principal components analysis can often mask the multidimensionality of items. When connected as a consequent factor to Compatibility, the multidimensionality of the items become apparent with the Usefulness items dominating the construct. The Ease of Use items in turn were more representative when the Compatibility construct was replaced with Result Demonstrability.

Thus, the results of this second empirical example demonstrate how second generation analytic methods such as PLS can provide additional information that are often lacking among first generation methods. Equipped with the results of a first generation components analysis, a researcher might conclude that the items for the Contrived? construct were appropriate. Furthermore, he or she might incorrectly decide to sum these items to create an overall score and in turn use it in a regression analysis. But in the case of the PLS analyses, we are able to simultaneously examine how well the items related to their respective constructs and the appropriateness of the theoretical model connecting these constructs.

This perspective, therefore, suggests that the "two step" approach to latent variable modeling where measures are developed first prior to their use in a theoretical model may not be efficient nor effective. Instead, it is argued that researchers consider a one step approach where both theory and measures are developed and tested simultaneously. Given the fact that the final test of the reliability and validity of a set of items must be made within a particular theory, why not perform the initial measurement development in the same context? To our knowledge, there has not been any IS studies adopting such a procedure.

In conclusion, we hope that researchers using an instrument that was either: 1) created in the traditional factor analytic procedure (i.e., without being embedded in a theoretical model) or 2) used in a different theoretical model will still check for appropriate reliability and validity in the holistic fashion presented in this paper. All to often, researchers use such instruments in a new theoretical model without checking to see whether it is generalizable to this new context. When they do, it is likely done in the atheoretical fashion of a principal components or factor analysis. We hope that we have shown that such instruments are not necessarily reasonable. Ideally, whether developed in the traditional manner or not, the appropriate test of reliability and validity of a set of items must be made simultaneously within the context of the theoretical model in which they are to be used. Finally, in terms of theory and measurement development, this paper suggests that a one step approach be considered rather than the prevalent two step approach. To test for this will more than likely involve the use of second generation analytical methods such as PLS.

Appendix

RELATIVE ADVANTAGE

VRA1 Using Voice Mail in my job enables (would enable) me to accomplish tasks more quickly.

VRA2 Using Voice Mail improves (would improve) my job performance.

EASE OF USE

VEOU1 Learning to operate Voice Mail is (would be) easy for me.

VEOU2 I find (would find) it easy to get Voice Mail to do what I want it to do.

COMPATIBILITY

VCPT1 Using Voice Mail is (would be) compatible with all aspects of my work.

VCPT2 Using Voice Mail is (would be) completely compatible with my current situation.

VCPT3 I think that using Voice Mail fits (would fit) well with the way I like to work.

VCPT4 Using Voice Mail fits (would fit) into my work style.

VOLUNTARINESS

VVLT1 My superiors expect (would expect) me to use Voice Mail.

VVLT2 My use of Voice Mail is (would be) voluntary (as opposed to required by my superiors or job description).

VVLT3 My boss does not require (would not require) me to use Voice Mail.

VVLT4 Although it might be helpful, using Voice Mail is certainly not (would not be) compulsory in my job.

RESULT DEMONSTRABILITY

VRD1 I would have no difficulty telling others about the results of using Voice Mail.

VRD2 I believe I could communicate to others the consequences of using Voice Mail.

VRD3 The results of using Voice Mail are apparent to me.

VRD4 I would have difficulty explaining why using Voice Mail may or may not be beneficial.

all of the preceding used a 7 point Likert scale
(extremely likely, quite, slightly, neither, slightly, quite, extremely unlikely)

INTENTION

VINT1 I presently intend to use Voice Mail regularly:

(7 pt scale using: extremely likely, quite, slightly, neither, slightly, quite, extremely unlikely)

VINT2 My actual intention to use Voice Mail regularly is:

(7 pt scale using: extremely strong, quite, slightly, neither, slightly, quite, extremely weak)

VINT3 Once again, to what extent do you at present intend to use Voice Mail regularly:

(11 pt scale from 0 to 10 with anchors Definite no, Definite Yes)

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, 227-247.

Barclay, D., Higgins, C., and Thompson, R. "The Partial Least Squares (PLS) Approach To Causal Modeling: Personal Computer Adoption And Use As An Illustration, Technology Studies, forthcoming.

Chin, W. W. & Gopal, A., "An Examination Of The Relative Importance Of Four Belief Constructs On The GSS Adoption Decision: A Comparison Of Four Methods," *Proceedings Of The Twenty-Sixth Annual Hawaii International Conference On System Sciences*, Vol. 4, 1993, pp. 548-557.

Davis, F. D. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), September 1989, 319-340.

Davis, F.D. "User Acceptance of Information Technology: System Characteristics, User Perceptions, and Behavioral Impacts," *International Journal of Man-Machine Studies*, 38, September 1993, 475-487.

Davis, F.D., Bagozzi, R. and Warshaw, P. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science* (35:8), August 1989, 982-1003.

Fornell, C. & Yi, Y. "Assumption of the two-step approach to latent variable modeling," *Sociological Methods & Research*, 20(3), 1992, 291-320.

Fornell, C. "The blending of theoretical and empirical knowledge in structural equations with unobservables." In *Theoretical Empiricism*, edited by H. Wold, New York: Paragon, 1989, 153-173.

Harman, H. H. *Modern Factor Analysis*. Third edition. Chicago: The University of Chicago Press, 1976.

Hendrickson, A.R., Massey, P.D. and T.P. Cronan. "On the Test-Retest Reliability of the Perceived Usefulness and Perceived Ease of Use Scales." *MIS Quarterly*, (17:2), 1993, 227-230.

Klenke, K., "Construct Measurement In Management Information Systems: A Review And Critique Of User Satisfaction And User Involvement Instruments," *INFOR*, 30(4), 1992, 325-348.

Mathieson, K. "Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior," *Information Systems Research*, (2:3), September 1991, 173-191.

Segars, A. H. and Grover, V. "Re-Examining Perceived Ease of Use and Usefulness: A Confirmatory Factor Analysis," *MIS Quarterly* (17:4), December 1993, 517-525.

Straub, D. W., "Validating Instruments In MIS Research," *MIS Quarterly*, June, 1989, 147-170.

Suppe, F. *The Structure Of Scientific Theories*. Second edition. Urbana, Il: Univeristy of Illinois Press, 1977.

Velicer, W. F. & Jackson, D. N. "Component analysis versus common factor analysis: Some further observations," *Multivariate Behavioral Research*, 25, 1990, 97-114.

Wold, H. "Systems under indirect observation using PLS. In C. Fornell (Ed.), *A Second Generation of Multivariate Analyis*, Volume 1: Methods, New York: Praeger, 1982, 325-347.

Zmud, R. W. & Boynton, A. C., "Survey Measures and Instruments in MIS: Inventory and Appraisal," in *The Information Systems Research Challenge: Survey Research Methods*, vol. 3, K. L. Kraemer (Ed.), Boston: Harvard Business School, 1991, 149-180.