

SCHOOL OF BUSINESS AND ECONOMICS
UNIVERSITY OF MINNESOTA, DULUTH
BUREAU OF BUSINESS AND ECONOMIC RESEARCH
WORKING PAPER NUMBER 77-1

RELIABILITY: A TUTORIAL NOTE

by

Jon L. Pierce*
L. L. Cummings**

*Assistant Professor of Organization Theory & Behavior, School of
Business and Economics, University of Minnesota-Duluth, Duluth,
Minnesota 55812

**Professor of Organization Behavior, Graduate School of Business,
University of Wisconsin, Madison, Wisconsin 54306

Working Paper: March, 1977

DO NOT CITE OR QUOTE WITHOUT PERMISSION #77-1

SCHOOL OF BUSINESS AND ECONOMICS
UNIVERSITY OF MINNESOTA, DULUTH
BUREAU OF BUSINESS AND ECONOMIC RESEARCH
WORKING PAPER NUMBER 77-1

RELIABILITY: A TUTORIAL NOTE

by

Jon L. Pierce*
L. L. Cummings**

*Assistant Professor of Organization Theory & Behavior, School of Business and Economics, University of Minnesota-Duluth, Duluth, Minnesota 55812

**Professor of Organization Behavior, Graduate School of Business, University of Wisconsin, Madison, Wisconsin 54306

Working Paper: March, 1977

ABSTRACT

The purpose of this tutorial note is to call attention the need for a more detailed examination of reliability in organizational behavior research. Reliability is a multi-dimensional phenomena for which there has been a general neglect in the examination and reporting of the stability coefficient for research measures. While there is the tendency to accept established measures as reliable and valid the psychometric properties of research instruments should be examined on a sample-by-sample basis since the validity of a measure is never definitive. The increased use of employee comparisons in inter-unit and inter-organizational research and longitudinal analysis make the examination of the stability and the internal consistency coefficient even more critical.

RELIABILITY: A TUTORIAL NOTE

One of the defining characteristics of scientific rigor in survey research on organizations is the reliability of the measures employed. Unfortunately, it seems to be assumed that the most frequently used and cited research instruments in organizational psychology generally possess demonstrated reliability.

While estimates of the internal consistency of a scale (i.e., the homogeneity of multiple item scales frequently operationalized as coefficient alpha) are usually reported, this minimal expectation of scientific analysis has become customary only during the past few years. It is typically the case that estimates of the stability of a research measure (i.e., the degree of correspondence between the administration and re-administration of a measure across time--test-retest reliability) are not available in the literature for most questionnaires upon which our "knowledge" is based. As a specific case in point, not a single study reported in the last two years (1975 and 1976) in The Academy of Management Journal and The Journal of Applied Psychology utilizing one of the two major indices of satisfaction (the Minnesota Satisfaction Questionnaire and the Job Descriptive Index) reported test-retest reliability estimates. In fact, the general case can be made that this deficiency retards or even prevents our ability to draw meaningful conclusions from the measurement of several of the most frequently referenced constructs in organizational behavior (e.g., leader behavior or style, organizational climate, and motivation when operationalized through expectancy constructs).

There are several reasons why this weakness (i.e., failure to examine and report scale reliability-internal consistency and stability-

coefficients on a sample-by-sample basis) in our measurement and/or our reporting procedures should be taken seriously. First, as research has begun to focus on inter-unit and inter-organizational comparisons and analyses, it becomes imperative to establish the psychometric properties (e.g., stability and internal consistency estimates) of the measuring instruments used in such studies. This is needed in order to evaluate the validity of such comparisons. Unfortunately, most studies presently reported in the literature comparing dependent variables across units or across organizations either do not report estimates of the psychometric characteristics of the instruments employed, or report estimates based only on total samples. That is, estimates are not reported for each unit across which comparisons are made.

Secondly, the interpretation of any differences in relationships between an independent and a dependent variable (e.g., a change in task design and subsequent satisfaction with the task) between units or organizations necessitate knowledge of the stability of the criterion. Interpretations of such differences in the effects of changes in an independent variable when only scale (i.e., internal consistency) reliabilities are known, even if acceptably high, is less accurate, less specific, and subject to much greater risks of inappropriate generalization.

Thirdly, it is generally known that differences in the effects of an independent variable (e.g., task design) on a dependent variable (e.g., satisfaction with the task) can be partially attributable to the differences in the measures used to assess the dependent variable across studies. When such a cause of discrepant findings is suspected, knowledge of the differential stabilities of the dependent variable measure is crucial.

Thus, there are several reasons to assess reliability (via different techniques) of variable indices when attempting to conduct comparative research in organizational behavior. For illustrative purposes only, this paper reports data comparing two forms of reliability (internal consistency and stability) of two satisfaction measures across three units of the same organization. These data illustrate the dangers of using reliability estimates from a single method, reliability estimates derived from total sample data, or reliability estimates based on normative data available in previously reported studies, to infer the reliability of a measuring instrument.

METHOD

Sample and Data Collection

The data, collected as a part of a larger study, was drawn from a sample of thirty employees in the home office of a multiple-line insurance company (5,000 employees). The predominately female sample (approximately 67 percent), with an average age of 34 years, was taken from three generically distinct work units (personnel, education and group claims). There was a one-month time interval between the two administrations of the research instrument.

Instruments

The short form of the Weiss, Dawis, England and Lofquist (1967) Minnesota Satisfaction Questionnaire (MSQ) was employed to measure general satisfaction (a twenty item scale), intrinsic satisfaction (twelve items) and extrinsic job facet satisfaction (six items). Kind of work satisfaction (a six item scale) was measured by Smith's (1976) Index of Organizational Reactions (IOR).

Analysis

Two estimates of reliability (stability and internal consistency) will be examined. Reliability will be assessed through test-retest correlation coefficients and Cronbach's (1951) coefficient alpha. With the individual serving as the unit of analysis, both reliability estimates will be examined for the entire sample as well as on a work unit-by-work unit basis.

RESULTS

Table 1 presents test-retest reliability estimates, means, standard deviations, and number of scale items for each of the affective response variables for the full sample.

The test-retest reliability for the MSQ short form was not reported by Weiss, et. al., (1967) in the development of their satisfaction measure. Instead the stability of the short form is based on their generalization from the MSQ long form, where a stability coefficient of .89 was obtained over a one-week period and .70 over a one-year period. The stability estimates for the three MSQ scales from this sample are questionably low (general satisfaction .53; intrinsic satisfaction .50; and extrinsic satisfaction .63). The test-retest correlation coefficient for the IOR's kind of work satisfaction was .90.

Internal consistency estimates for each satisfaction scale at time one and time two are shown in Table 2. The internal consistency coefficient for the IOR and the general measure of satisfaction were greater than .75. Only one measure of intrinsic satisfaction and four of the extrinsic satisfaction scales have internal consistencies lower than .70. While standards for acceptable estimates of internal consistency are

TABLE 1

Test-Retest & Descriptive Statistics (n=30)

Variable	Scale Items	Mean		Standard Deviation		Test-Retest Coefficient
		T ₁	T ₂	T ₁	T ₂	
MSQ:						
General Satisfaction	20	72.2	74.3	13.1	10.3	.53
Intrinsic Satisfaction	12	46.4	47.4	8.1	6.5	.50
Extrinsic Satisfaction	6	18.4	19.5	5.3	4.0	.63
IOR:						
Kind of Work Satisfaction	6	22.4	22.7	4.6	4.0	.90

TABLE 2

Internal Consistency

Variable	Total Sample (n=30)		Unit 1 (n=5)		Unit 2 (n=12)		Unit 3 (n=10)	
	Time		Time		Time		Time	
	1	2	1	2	1	2	1	2
MSQ:								
General Satisfaction	91	89	78	91	93	92	93	81
Intrinsic Satisfaction	88	85	59	80	89	88	93	71
Extrinsic Satisfaction	86	68	67	84	88	63	88	67
IOR:								
Kind of Work Satisfaction	92	86	84	84	92	82	93	89

NOTE: (a) Decimals have been omitted.

(b) An examination of the internal consistency estimates for the total sample was based on a sample of n=30. The examination of this reliability estimate for each work unit is based on an n=27 since three respondents were unidentifiable in terms of their respective work units.

admittedly tenuous, Nunnally (1967, p. 226) suggests that alpha values $\geq .70$ are necessary for established research scales and that values of $\geq .50$'s and $\geq .60$'s suffice for measures in the early stage of development. One alpha value for unit two (.67), unit three (.67), and for the sample of 30 (.68) are judged low in accord with this standard of $\geq .70$, while two of the alpha values (.59, .67) for unit one are lower than this standard. In accord with current research practice, however, these reliability values would generally be judged acceptable.

Stability correlation coefficients (i.e., test-retest reliabilities) for each of the three units are presented, along with their respective means and standard deviations, in Table 3. The stability coefficients for the three MSQ scales and the IOR are $> .80$ and are judged respectable for two of the three work units. In work unit 3 the MSQ satisfaction measures ranged from $-.06$ (for general satisfaction) to $.30$ (extrinsic satisfaction), while the IOR's test-retest coefficient was $.85$ for the same respondents.

The absolute value of the mean difference scores (time one minus time two scores) are shown in Table 4. For each of the three MSQ scales, the largest shifts occurred, as noted by the low test-retest reliability estimates identified in Table 3, in work unit three. Examination of the difference scores (d) on a respondent-by-respondent basis reveals a number of sizable changes between the two administrations. For the general satisfaction measure two of the ten respondents had difference scores above their work unit mean. One respondent had a 58 point general satisfaction differential across the one-month time period. One respondent had a difference score greater than the mean for intrinsic satisfaction

TABLE 3

Work Unit Test-Retest Reliabilities

Variable	Work Unit 1					Work Unit 2					Work Unit 3				
	Stblty (n=5)	Mean		Std Dev		Stblty (n=12)	Mean		Std Dev		Stblty (n=10)	Mean		Std Dev	
		1	2	1	2		1	2	1	2		1	2		
MSQ:															
General Satisfaction	.94	70.2	67.8	8.3	11.3	.91	73	73.4	13.8	11	-.06	70.6	77.6	16	8.3
Intrinsic Satisfaction	.80	47	44.2	4.6	5.2	.87	46.2	46	8.0	7.7	.04	45.7	50.2	10.7	4.4
Extrinsic Satisfaction	.93	17.6	18	3.8	4.9	.86	18.7	19.3	5.7	3.9	.30	17.5	19.7	6.2	4.4
IOR:															
Kind of Work Satisfaction	.92	20.6	20.6	4.0	3.5	.89	21.4	21.4	4.8	3.6	.85	24.1	24.8	4.7	4.0

TABLE 4
Time 1 - Time 2 Difference Scores

Variable	Total Sample		Unit 1 (n=5)		Unit 2 (n=12)		Unit 3 (n=10)	
	Average	Range	Average	Range	Average	Range	Average	Range
MSQ:								
General	5.5	-58 + 9	2.7	-1 + 9	4.3	-15 + 7	9.2	-58 + 5
Intrinsic	3.8	-36 + 8	2.3	0 + 8	2.9	-9 + 6	5.9	-36 + 3
Extrinsic	2.5	-19 + 6	1.3	-2 + 3	2.5	-4 + 6	3.6	-19 + 2
IOR:								
Kind of Work	1.5	-6 + 4	1.2	-2 + 2	1.7	-4 + 4	1.7	-6 + 3

($d = 36$ points) and two had a difference score greater than the extrinsic satisfaction mean ($d = 19$ and 6).

DISCUSSION

The scientific rigor of survey research on organization is in part defined by reliability of the research methods employed. Reliability estimates of measurement in contemporary organizational psychology essentially derives from estimates of internal consistency (e.g., coefficient alpha). Many researchers have tended to assume demonstrated reliability for the more popular research instruments; consequently, there has been the frequent tendency to ignore the fact that "validation is an unending process" (Nunnally, 1967, p. 75).

During the past few years the reporting of estimates of internal consistency has become standard while an examination of the stability coefficient has been neglected. As demonstrated in this research note the adoption of a reliability determination standard based solely on estimates of internal consistency can be misleading. In accord with that practice the data from this investigation would have gone unchallenged due to: (a) the a priori reputation of the research instrument as a criterion measure; and (b) the estimates of internal consistency (MSQ, time 1, $n = 30$; coefficient alphas $.86 - .91$).

The data employed here, for illustrative purposes only, highlight a number of tutorial perspectives. The reliance on the a priori demonstrated reliability of a research instrument could be misleading as reliability estimates alone provide an incomplete perspective on the overall reliability of a research instrument. The data examined here reveal

acceptable levels of internal consistency (time one: .86 - .92) and questionably low stability coefficients (.50 - .90). The reliance on reliability estimates based on the total sample, when the focus of the inquiry is on inter-unit or inter-organizational research, is also a questionable procedure thus challenging the validity of comparative analysis. The sample data illustrated variability in internal consistency estimates, as well as differentials in the stability coefficient across the multiple analytic units. The researcher should entertain the consideration of eliminating the data from that unit for which low reliability estimates on the criterion measure were found. Finally, the use of multiple criterion variables is a sound research practice. In unit three, weak stability coefficients were found for only one of the two criterion measures. This tends to suggest that the subjects had response problems with one of the two criterion measures and that reliable data could be obtained from the same group of respondents.

While the data presented here is limited by the size of the sample, it is not purported that the data has merit in terms of new empirical findings. Rather the value of the data is embedded in its tutorial role which highlighted three major problems. The recommendations for future research are embedded in the following three positions. First, whether the researcher is engaged in single or multiple unit investigations, there is a need for a more complete examination of the reliability of the research instruments employed. Acceptance of demonstrated reliability estimates and the examination of internal consistency is not adequate. Secondly, as researchers increasingly make use of inter-unit and inter-organizational comparisons the validity of such comparisons cannot

be established without an examination of the reliability of the measures on a unit-by-unit basis. Thirdly, with the use of longitudinal analysis, knowledge of the stability of research instruments becomes critical.

REFERENCES

1. Cronback, L. J. "Coefficient Alpha and the Internal Structure of Tests," Psychometric, 1951, 16, 297-334.
2. Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, 1967.
3. Smith, F. J. "Index of Organization Reactions (IOR)" Catalog of Selected Documents in Psychology, 6, 1976 (May), Manuscript #1265. American Psychological Association.
4. Weiss, D. J., Dawis, R. V., England, G. W., and Lofquist, L. H. Manual for the Minnesota Satisfaction Questionnaire. Minneapolis: University of Minnesota, Industrial Relations Center, Work Adjustment Project, 1967.