Human Relations, Vol. 45, No. 10, 1992

Reply Article

Replication Requisites: A Second Look at Klenke-Hamel and Mathieu (1990)

Arthur G. Bedeian,[1,4] Kevin W. Mossholder,[1] Edward R. Kemery,[2] and Achilles A. Armenakis[3]

In a recent article, Klenke-Hamel and Mathieu (1990) report an operational replication and extension of the Bedeian and Armenakis (1981) investigation of the effects of role conflict and role ambiguity on a number of job-related attitudes. Utilizing multiple samples involving diverse jobs, they tested the generalizability of the Bedeian and Armenakis framework and explored the effects of employee personal characteristics on its focal variables. Although we encourage replication research, the Klenke-Hamel and Mathieu study contains concerns which may render its findings suspect. We examine some of these concerns, indicating where we feel interpretational problems surface, and comment on the role of replication in extending theory.

KEY WORDS: role ambiguity; role conflict; replication.

INTRODUCTION

Replication is a basic requirement for encouraging confidence in the internal validity of studies, as well as the reliability and generalizability of research findings (Campbell & Jackson, 1979). While, as typically defined, replications are studies that provide additional tests of an already tested research hypothesis (Cooper & Rosenthal, 1980), they may vary widely in the closeness with which they duplicate a target study. *Literal* or *exact* replications duplicate procedures as precisely as possible. By contrast, *operational* replications stress the duplication of a target study's proce-

[1]Department of Management, Louisiana State University, Baton Rouge, Louisiana 70803-6312.
[2]Department of Management, University of Baltimore, 1420 N. Charles Street, Baltimore, Maryland 21201-5779.
[3]Department of Management, Auburn University, Auburn, Alabama 36830-5241.
[4]Requests for reprints should be addressed to Arthur G. Bedeian, Department of Management, Louisiana State University, Baton Rouge, Louisiana 70803-6312.

0018-7267/92/1000-1093$06.50/1

random measurement error in social science constructs, some degree c incongruity or slippage is expected (Dawis, 1987). Moreover, for reasor both theoretical and practical, operational replications may measure model's constructs in ways differing from a target study. However, as Jame Mulaik, and Brett (1982) note, the degree of slippage allowed in confi matory analysis is smaller than researchers enjoy in exploratory analysis (1 58). Indeed, when the slippage is too much at variance from a target stud an intended confirmatory study is no longer an operational replication, bu a test of a different conceptual model.

Assessments concerning correspondence between a construct and i respective measure are invariably shaped by many factors. With this i mind, it is instructive to examine the Klenke-Hamel and Mathieu study. provides examples of problems of congruity in both construct labeling an measurement. As an example of the former, Klenke-Hamel and Mathie identify role conflict and ambiguity as two facets of role strain. By contras the Bedeian and Armenakis model treats role conflict and ambiguity i standard fashion as facets of perceived role stress (cf. Schneider & Bower 1985). This difference in labeling creates the potential for misunderstandin when examining constructs across studies and may confuse the status c the exogenous and endogenous variables being investigated. Stress an strain are old terms with established meanings and are not synonymou (Kahn, 1981, p.54). When stress is a cause, strain is a resulting outcome Construct-derived variables must be labeled in an operationally specifi manner to avoid such slippage.

The concern for congruity in construct measurement centers o relationships between operational measures and the particular construct they represent. Items included in a measure should sufficiently, though par simoniously, sample the domain of a target construct. At the same time construct definition in a target model should be robust enough so that it fit does not depend on a specific set of measures. Otherwise a model run the risk of being measure (i.e., method) dependent.

If one of several measures of the same construct inadequately sample its content domain, its explanatory power may be diminished (or eve spuriously enhanced) in comparison to the other measures. As regards th Klenke-Hamel and Mathieu study, item sampling adequacy may have bee a problem as it relates to the measurement of role conflict and ambiguity Both variables were gauged by two items extracted from measures used b Caplan, Cobb, French, Harrison, and Pinneau (1975). Whether these iten pairs adequately cover each construct's intended content is at least ques tionable. To the extent that they are content deficient, tests of model involving these constructs will be limited in generalizability. The Rizzo House, and Lirtzman (1970) role perception scales used by Bedeian an

Armenakis sample a wider content universe by encompassing multiple kinds of role conflict (i.e., intrarole conflict, interrole conflict, intersender conflict) and ambiguity (i.e., role predictability, role clarity, role certainty), using six and eight items, respectively. These scales were specifically chosen for their established psychometric properties (Schuler, Aldag, & Brief, 1977). We note that though more used, even these scales are not free from construct validity criticisms (Netermeyer et al., 1990).

A related problem also surfaces with regard to Klenke-Hamel and Mathieu's index of tension. The problem with construct measurement here is not content deficiency but contamination. The Bedeian and Armenakis model was constructed to consider stress–strain relationships occurring in a work context. Thus, the measures used focused on the work domain. Regarding the tension construct in the Bedeian and Armenakis model, job-related tension is the extent to which an individual feels bothered by job-related matters. Bedeian and Armenakis used a nine-item scale developed by Lyons (1971) to tap this construct. It was specifically constructed to determine the frequency with which respondents report feeling bothered by a variety of work-related factors. Exemplary items include: "Not knowing what your immediate superior thinks of you, how he/she evaluates your performance" and "Feeling that you have to do things on the job that are against your better judgment." The counterpart to the Lyons scale in the Klenke-Hamel and Mathieu study was a 20-item version of Goldberg's General Mental Health Questionnaire (GMHQ). The GMHQ was developed specifically to detect psychiatric disturbance (Goldberg, 1971, p. 1). An inspection of GMHQ items suggests that it is best described as a measure of overall mental health rather than an index of job-related tension. Exemplary items include: "Have you recently been getting out of the house as much as usual?" and "Have you recently found at times you couldn't do anything because your nerves were so bad?" While psychiatric illness may subsume some of the effects of job-related tension, it obviously taps a wider domain. Thus, the possibility of construct slippage must be considered to the extent to which these constructs do not overlap.

Klenke-Hamel and Mathieu acknowledge that some inconsistencies between their results and earlier tests of the Bedeian and Armenakis model are likely due to measurement differences. However, this acknowledgment, by itself, neither rectifies nor addresses validity as a criterion for assessing the GMHQ's utility in measuring job-related tension. The authors do contend (on a *post hoc* basis) that the GMHQ appears to be a reasonable measure for the Bedeian and Armenakis framework because it yielded significant results. Such *post hoc* reasoning is not a good substitute for *a priori* theory because hidden biases, unrelated to theoretical

considerations, may account for observed linkages (Bagozzi, Yi, & Phillips, 1991). Unless a measure's construct validity (i.e., the degree to which it captures the construct it is intended to gauge) can be established prior to its use in a model, its value in building a body of knowledge must remain suspect.

SAMPLING ISSUES

A stated purpose of the Klenke-Hamel and Mathieu study was to test the Bedeian and Armenakis model with a wider variety of employee populations, implying the intent to determine if the model's hypothesized relationships could be generalized to these populations. With a heterogeneous sampling model (Cook, Campbell, & Peracchio, 1990), each time a study is successfully replicated (and the sample of individuals and settings fits within pre-defined boundary conditions), confidence in the generalizability of results increases. Generalizing findings in this manner has been referred to as "triangulation" by Sackett and Larson (1990), and stands in contrast to the usually unpracticable procedure of using statistical sampling to generalize to an entire population, in this case working individuals. Tradeoffs between generalizability of a model and the internal validity of tests of the model in different samples often occur (Cook et al., 1990) and are typically decided on the basis of researcher judgment. From our viewpoint, the potential generalizability gains offered by the Klenke-Hamel and Mathieu study may be offset by problems associated with sampling diversity.

It is desirable that measures used across various samples possess sufficiently homogeneous psychometric characteristics. If not, their meaning may actually be different depending on the sample. Some of the measures used by Klenke-Hamel and Mathieu, notably the two-item measures, display varying levels of internal consistency, as gauged by the correlations between the constituent items. Specifically, there are significant differences (chi-square test for correlational differences; see Hays, 1973, p. 664) across samples in the between-item correlations for role ambiguity ($p < .001$), role conflict ($p < .06$), and propensity to leave ($p < .001$). Apparently, the items comprising these measures were interpreted differently by respondents in the various samples, thus raising questions about reliability across the samples (cf. Dawis, 1987).

An additional concern in sample framing is attempting to explain findings of model variables that may have been influenced by the composition of a sample. This goes beyond the issue of sample representativeness (Sackett & Larson, 1990) and speaks more to the internal validity of a replication across samples. In their extensions of the Bedeian

and Armenakis model, Klenke-Hamel and Mathieu introduced employee personal characteristics (gender, education, age, and marital status) into the target model and found that these improved the fit depending on the specific sample and personal characteristic. It appears that the samples were roughly similar with respect to age and tenure, but exhibited more variation with regard to educational level, gender composition, and percentage married.

As noted above, testing the Bedeian and Armenakis model across diverse samples has the potential to increase the generalizability of results. However, the original Bedeian and Armenakis model did not include hypotheses about personal characteristics. When such variables are included, as in the model's extension, concern for possible interactions with sampling become salient since the composition of the samples with regard to personal characteristics could influence results in unknown ways.

The most extreme example of this problem is in connection with gender. For the blue-collar sample, Klenke-Hamel and Mathieu concluded that women report greater propensity to leave, tension, and job satisfaction than their male counterparts. The efficacy of this finding, however, is questionable given that this sample was comprised of some 184 women and only three men. In the faculty sample, no males were included while in the engineering sample, 91% were male. Inconsistencies in gender composition across samples render it difficult to determine the role gender might play in the Bedeian and Armenakis model. Though not as stark, similar inconsistencies exist for education (cf. faculty vs. blue-collar samples) and marital status (cf. blue-collar vs. the other three samples).

Range restriction problems within samples or selection-by-treatment interactions across samples (cf. Cook et al., 1990) may unduly influence results. For example, disparate splits (i.e., unequal *n*s) on dichotomous variables may induce different observable relationships for reasons unrelated to substantive concerns (Kemery, Dunlap, & Bedeian, 1989). Researchers must exercise caution to insure that when extending a theoretical model, the variables they introduce are testable within the samples collected. This is especially true when new variables are added on a *post hoc* basis and "convenience" samples (Sackett & Larson, 1990) are used. In such instances, possible interactions between samples and new variables would not have entered into sampling considerations.

In sum, we are in agreement with Klenke-Hamel and Mathieu's conclusion, as well as that of others (e.g., Netermeyer et al., 1990), that role stress likely influences different employee populations in unique ways. With respect to the Klenke-Hamel and Mathieu study, however, questions about proper sample framing must be raised to place their findings in perspective.

THEORY TRIMMING AND MODEL COMPARISONS

Cooper and Rosenthal (1980) distinguish between replication and extension studies. They view extension studies as moving beyond an original hypothesis to contribute findings undetected in previous studies. Klenke-Hamel and Mathieu also offer an extension of the Bedeian and Armenakis model by exploring alternative models. They adopt a three-phase procedure for this purpose. First, they trimmed nonsignificant paths from the Bedeian and Armenakis model from each sample. Second, omitted parameter tests were performed for each of the three endogenous variables (i.e., propensity to leave, tension, and job satisfaction) in the Bedeian and Armenakis model in order to identify any further significant effects. Finally, employee personal characteristics (i.e., gender, education, age) were introduced to explain additional sources of influence. Klenke-Hamel and Mathieu clearly acknowledge that these tests are strictly exploratory being particularly susceptible to chance correlations.

This aside, certain questions remain. The first and most direct is that Klenke-Hamel and Mathieu offer no theoretical rationale for their selection of employee characteristics and the explicit proposed causal relationships. Such a rationale constitutes what Whetten (1989) terms the "theoretical glue" that holds a model together. The central question of why credence should be given to the proposed revised model incorporating employee characteristics is unaddressed. Adding a new variable to an existing model may not itself constitute a theoretical contribution in the absence of an argument specifying the rationale under which the variable is introduced.

Having acknowledged the danger of spurious relationships resulting from empirically-derived models, an additional point should be mentioned. This concerns the act of *simultaneously* trimming and adding paths to a model. It is difficult to specify what such a resulting model means for disconfirming a target model. The ambiguity resulting from such extension studies would seem especially great, particularly if they are unsuccessful in replicating a target model. In addition to the uncertainty associated with operational replications in general, failure to confirm a target model that has been simultaneously trimmed and extended may mean that its supporting theory was incorrect, that the proposed extension represented by the new paths was in error, that the measurement instruments chosen to test the new model were inappropriate, and so on.

One way of avoiding such ambiguity is to employ what Sidman (1960) labels the "baseline" technique of systematic replication, whereby a study is replicated both operationally and constructively. This technique is essentially identical to Rosenthal's (1979, 1990) suggestion that a "replication battery" be employed to determine if a failure to replicate is due to the

nonreplicability of a target study or to the inexactness of replication procedures. In its most basic form, a replication battery requires two replications. One, the operational replication, is as similar as possible (though not identical) to an original study. The other, the constructive replication, is somewhat dissimilar in ways pre-designed and theoretically directed. If both replications yield consistent results in terms of effect sizes, the reliability of an original study's results would be bolstered and its generalizability expanded to include the boundary conditions of the constructive replication.

Finally, the Klenke-Hamel and Mathieu study raises an intriguing question. In particular, how does one compare causal models across different samples? There is no general answer to this question, and no clear-cut standards exist for comparing model fit in multiple samples. Klenke-Hamel and Mathieu employ Hotchkiss's (1976) technique for comparisons across samples. In brief, this approach permits simultaneous comparisons between independent variables within and between samples, as when one wishes to compare the magnitudes of coefficients associated with two or more independent variables. The overall fit of a target model across samples is assessed by means of the generalized multiple correlation coefficient (Q). The closer Q is to 1.0, the better a model's goodness-of-fit.

As indicated by nearly identical Q-values, the Bedeian and Armenakis target model fits nearly equally well in three of Klenke-Hamel and Mathieu's samples (staff $Q = .939$, $\chi^2 = 16.002$, $p > .05$; engineering $Q = .903$, $\chi^2 = 7.713$, $p > .05$; faculty $Q = .912$, $\chi^2 = 5.170$, $p > .05$) implying that the model is strongly supported by the data for all but the blue-collar sample ($Q = .773$, $\chi^2 = 43.98$, $p < .001$). However, a difficulty arises in situations where the overall fit of a target model is roughly equivalent across samples, but each sample differs in the number of statistically significant paths. In the Klenke-Hamel and Mathieu (1990) study, seven, five, and three of eight paths were significant, respectively, for the staff, engineering, and faculty samples. How this result should be interpreted in light of the similar Q-values presents a dilemma that can arise when testing causal models.

Implied in any theoretical model is a set of functional equations which permit the assessment of theoretical linkages within the model. From a multiple regression standpoint, each endogenous variable contained in a causal model is actually a criterion variable which is regressed on all antecedent variables to obtain estimates of path coefficients. These coefficients represent evidence of direct and indirect effects within the context of an hypothesized framework. Statistical evidence must be found for predicted linkages to be supported.

In this regard, Klenke-Hamel and Mathieu inspected individual target model linkages for statistical significance. The number of nonsignificant

paths in each analysis ranged from one in their staff sample to five in their faculty sample. Because of the number of nonsignificant paths observed, Klenke-Hamel and Mathieu concluded that the Bedeian and Armenakis model received support for only the staff sample. Examining the path coefficient results in light of sampling considerations suggests an alternative explanation based on statistical power concerns. The sample sizes for all four groups examined ranged from fairly small (69) to moderate (272), respectively, for faculty and staff. As size decreases, the probability of detecting linkage-based effects also decreases.

To demonstrate the effect of statistical power within the context of bivariate correlations, we assessed the statistical power to detect a population correlation of 0.3, when individual variables are measured with a reliability of 0.8 and an alpha of .05. The results of these analyses indicated that the estimated power for the faculty sample was .51, about a 50–50 chance of finding a statistically significant effect. The power statistics for the remaining samples were .64, .91, .98, for engineers, blue-collar, and staff, respectively. In support of a limited power interpretation for the pattern of nonsignificant paths, the number of such paths observed in the Klenke-Hamel and Mathieu study varied perfectly with the estimated power (and sample size) of each analysis. That is, in the sample with the least power (i.e., faculty), only three of eight linkages were supported, while the sample with the most power (i.e., staff) seven of eight linkages were supported. This suggests that some of the sample-based differences in model support may be due to differences in statistical power.

When considering statistical power (as well as Type I and II errors), it is reasonable to rely on an omnibus test to assess model fit. Hotchkiss (1976) uses the Q statistic. Structural equation modeling (SEM) packages (e.g., LISREL VII; Jöreskog & Sörbom, 1987) also provide omnibus measures. Research sophistication in this area is increasing, and as a result, there are a host of goodness-of-fit measures which, while differing in their assumptions, permit one to infer degrees of model support (Bentler, 1990). SEM also permits the simultaneous assessment of model fit across samples, a procedure that is especially suited for replication efforts. A full discussion of these procedures is beyond the scope of this paper. The interested reader is referred to Bollen (1989) for examples of how SEM may be used in this context.

CONCLUSION

The issues discussed in this paper are obviously not unique to the particular studies or models examined. We have been faced with these and similar issues while investigating other researchers' models and have ex-

perienced varying degrees of success in avoiding replication pitfalls in our own work. The complexities and requirements associated with causal modeling are many; cautions raised here echo what has been said in a more general context (e.g., James et al., 1982). Examining such cautions in a specific context (i.e., the Klenke-Hamel and Mathieu study) provided a point of reference from which to consider their actual impact and reinforce an appreciation of the demands of research involving causal models.

Hopefully, these comments will be useful to others contemplating any of the various forms of replication discussed. As Rosenthal (1990, p. 15) notes, "replications are possible only in a relative sense." The same study can never be repeated by a different researcher. Indeed, the same study cannot even be repeated by the same researcher. At the very least, the subjects and researchers themselves become different individuals over time.

Finally, the extent to which an hypothesized model is consistent with obtained data does not indicate unequivocal support, as the observed data may also support other network orderings. In this regard, model fit is a necessary, but insufficient condition for concluding that a theory has been corroborated. In the last analysis, the adequacy of a model can only be determined by the soundness of the theoretical rationale on which it is based.

ACKNOWLEDGMENT

We thank Richard G. Netermeyer for vetting an earlier draft of this manuscript.

REFERENCES

BACHARACH, S. B. Organizational theories: Some criteria for evaluation. *Academy of Management Review,* 1989, *14,* 496-515.

BAGOZZI, R. P., YI, Y., & PHILLIPS, L. W. Assessing construct validity in organizational research. *Administrative Science Quarterly,* 1991, *36,* 421-458.

BEDEIAN, A. G., & ARMENAKIS, A. A. A path-analytic study of the consequences of role conflict and ambiguity. *Academy of Management Journal,* 1981, *24,* 417-424.

BENTLER, P. M. Comparative fit indexes in structural models. *Psychological Bulletin,* 1990, *107,* 238-246.

BOLLEN, K. A. *Structural equations with latent variables.* New York: Wiley, 1989.

CAMPBELL, K. E., & JACKSON, T. T. The role of and need for replication research in social psychology. *Replications in Social Psychology,* 1979, *1*(1), 3-14.

CAPLAN, R. D., COBB, S., FRENCH, J.R.P., JR., HARRISON, R. V., & PINNEAU, S. R., JR. *Job demands and worker health* (HEW Publications No. NIOSH-75-160). Washington, D.C.: National Institute of Occupational Safety and Health, 1975.

COOK, T. D., CAMPBELL, D. T., & PERACCHIO, L. (1990). Quasi experimentation. In M. D. Dunnette and L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1). Palo Alto, CA: Consulting Psychologist Press, pp. 491-576.

COOPER, H. H., & ROSENTHAL, R. Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin,* 1980, *87,* 442-449.

DAWIS, R. V. Scale construction. *Journal of Counseling Psychology,* 1987, *34,* 481-489.

GOLDBERG, D. P. *The detection of psychiatric illness by questionnaire.* Oxford: Oxford University Press, 1971.

HAYS, W. L. *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart and Winston, 1973.

HOTCHKISS, K. A technique for comparing path models between subgroups: Standardized path-regression coefficients. *Sociological Methods & Research,* 1976, *5,* 53-76.

JAMES, L. R., MULAIK, S. A., & BRETT, J. M. *Causal analysis: Assumptions, models, and data.* Beverly Hills. CA: Sage, 1982.

JÖRESKOG, R. G., & SÖRBOM, D. *LISREL: Analysis of linear structural relations by the method of maximum likelihood* (Version VII). Chicago: National Education Resources, 1987.

KAHN, R. L. *Work and health.* New York: Wiley, 1981.

KEMERY, E. R., DUNLAP, W. P., & BEDEIAN, A. G. The employee separation process: Criterion-related issues associated with tenure and turnover. *Journal of Management,* 1989, *5,* 417-424.

KLENKE-HAMEL, K. E., & MATHIEU, J. E. Role strains, tensions, and job satisfaction influences on employees' propensity to leave: A multi-sample replication and extension. *Human Relations,* 1990, *43,* 791-807.

LYKKEN, D. T. Statistical significance in psychological research. *Psychological Bulletin,* 1968, *70,* 151-159.

LYONS, T. F. Role clarity, need for clarity, satisfaction, tension, and withdrawal. *Organizational Behavior and Human Performance,* 1971, *6,* 99-110.

NETERMEYER, R. G., JOHNSTON, M. W., & BURTON, S. Analysis of role conflict and role ambiguity in a structural equations framework. *Journal of Applied Psychology,* 1990, *75,* 148-157.

RIZZO, J. R., HOUSE, R. J., & LIRTZMAN, S. I. Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly,* 1970, *15,* 150-163.

ROSENTHAL, R. Replication in behavioral research. *Journal of Social Behavior and Personality,* 1990, *5*(4), 1-35.

ROSENTHAL, R. Replications and their relative utilities. *Replications in Social Psychology,* 1979, *1*(1), 15-23.

SACKETT, P. R., & LARSON, J. R., JR. Research strategies and tactics in industrial and organizational psychology. In M. D. Dunnette and L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1). Palo Alto, CA: Consulting Psychologist Press, 1990, pp. 419-490.

SCHNEIDER, B., & BOWEN, D. E. Employee and customer perceptions of service in banks: Replication and extension. *Journal of Applied Psychology,* 1985, *70,* 423-433.

SCHULER, R. S., ALDAG, R. J., & BRIEF, A. P. Role conflict and ambiguity: A scale analysis. *Organizational Behavior and Human Performance,* 1977, *20,* 111-128.

SIDMAN, M. Tactics of scientific research. New York: Basic, 1960.

WHETTEN, D. A. What constitutes a theoretical contribution? *Academy of Management Review,* 1989, *14,* 490-495.

BIOGRAPHICAL NOTES

ARTHUR G. BEDEIAN is the Ralph and Kacoo Olinde Professor of Management, Louisiana State University. He earned his DBA from Mississippi State University.

KEVIN W. MOSSHOLDER is the H. Norman Saurage, Jr./Community Coffee Endowed Professor of Management, Louisiana State University. He earned his PhD from the University of Tennessee.

EDWARD R. KEMERY is an Associate Professor of Management at the University of Baltimore. He earned his PhD from Auburn University.

ACHILLES A. ARMENAKIS is a Professor of Management at Auburn University. He earned his DBA from Mississippi State University.