Journal of Applied Psychology
1990, Vol. 75, No. 6, 738-742

0021-9010/90/$00.75

Item Content Versus Item Wording: Disentangling Role Conflict and Role Ambiguity

E. Kevin Kelloway and Julian Barling
Queen's University, Kingston, Ontario, Canada

Rizzo, House, and Lirtzman's (1970) Role Ambiguity Scale and Role Conflict Scales assess ambiguity with 6 negatively worded items and conflict with 8 positively worded items, respectively. This methodological confound between item wording and content precludes unambiguous interpretation. In the present study, confirmatory factor analysis of these 2 scales and Beehr, Walsh, and Taber's (1976) Role Overload Scale (which has positively and negatively worded items) was used to disentangle this confound. Across 2 independent samples ($N = 767$ and $N = 363$), a 3-factor model consistent with conceptual definitions of role ambiguity, conflict, and overload fit the data better than models with (a) one general role-stress factor, (b) a general role-stress and a method (item wording) factor, or (c) two method (positive and negative wording) factors. These results support the construct validity of Rizzo et al.'s (1970) scales; the consistency of the results across 2 independent samples suggests their generalizability.

Rizzo, House, and Lirtzman's (1970) Role Ambiguity Scale and Role Conflict Scale have been widely used in organizational psychology research. A vast amount of psychometric information is available for these scales (for a review, see Cook, Hepworth, Wall, & Warr, 1981), and at least two meta-analytic studies summarize the correlates of role conflict and role ambiguity (Fisher & Gitelson, 1983; Jackson & Schuler, 1985). Despite their widespread use, concern over the construct validity of Rizzo et al.'s (1970) scales continues unabated and has culminated recently in a call for a moratorium on their use (McGee, Ferguson, & Seers, 1989).

The debate on the construct validity of Rizzo et al.'s (1970) scales has been focused on the possibility that the two subscales reflect a single underlying construct, namely, role stress (Tracy & Johnson, 1981). This possibility emerges because the six items comprising the ambiguity scale reflect the *absence* of ambiguity. In contrast, the eight conflict items reflect the *presence* of role conflict. Thus, the substantive interpretation of these two scales is perfectly confounded with the direction of item wording, rendering any interpretation open to plausible, rival hypotheses.

In a recent confirmatory factor analysis of the role conflict and role ambiguity scales, McGee et al. (1989) contrasted the goodness of fit of a one-factor (role stress) model, a two-factor (role conflict and role ambiguity) model, and another two-factor model (general role stress together with a method factor on which all the negatively worded items loaded).[1] McGee et al. suggested that the last model provided the best fit to the data and concluded that the confound of item wording with item content compromises the construct validity of the scales, thereby precluding any substantive interpretations. However, McGee et al. could not provide unambiguous support for a model positing a general and a method factor because the direction of item wording (positive or negative) is perfectly confounded with item content (conflict and ambiguity). Consequently, it remains unclear whether McGee et al.'s method factor is a function of item wording or item content.

Accordingly, in the present study, we attempted to disentangle this confound through confirmatory factor analysis of three scales: Rizzo et al.'s (1970) Role Conflict Scale and Role Ambiguity Scale and Beehr, Walsh, and Taber's (1976) Role Overload Scale. The Role Overload Scale was included in our analysis for two reasons. First, like role conflict and ambiguity, role overload is a central component of role stress (Kahn, Wolfe, Quinn, Snoek, & Rosenthal, 1964). Second, the Role Overload Scale comprises two positively worded items and one negatively worded item, making it possible to disentangle item content from item wording. To do so, we empirically contrasted the goodness of fit of four plausible models by using confirmatory factor analysis. The first model represents a general role-stress

This research was supported in part by a doctoral fellowship (Grant 452-89-2138) from the Social Sciences and Humanities Research Council of Canada to E. Kevin Kelloway and by Social Sciences and Humanities Research Council of Canada Research Grant 410-88-0891 to Julian Barling.

We express appreciation to W. H. Cooper, Clive Fullagar, Karyl MacEwen, and two anonymous reviewers for constructive comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Julian Barling, Department of Psychology, Queen's University, Kingston, Ontario, Canada K7L 3N6.

[1] McGee, Ferguson, and Steers (1989) referred to this as a second-order model, comprising a single role-stress factor and a second-order method factor. We believe that their use of the term *second-order factor*, although correct, is potentially misleading. Although it is possible to operationalize second-order factors in terms of the observed variables (e.g., Schmid & Leiman, 1957), a second-order factor analysis commonly refers to the extraction of factors based on the analysis of covariances among first-order factors. In the current study, we followed McGee et al.'s (1989) methodology (e.g., in positing a two-factor model comparable to their second-order model), but we avoid reference to second-order factors in the interests of clarity.

dimension (Tracy & Johnson, 1981) and is reflected by one factor on which all items were hypothesized to load. The second model is consistent with the conceptual interpretation of role overload, conflict, and ambiguity as separate but related stressors and is reflected in a model specifying three oblique factors. The third model is consistent with the notion that there is a substantive role-stress factor and a methodological confound between item wording and item content (McGee et al., 1989). We posited a two-factor model comprising a general role-stress factor and a second, orthogonal, factor comprising all seven negatively worded items (i.e., the six role ambiguity items *and* the one negatively worded role overload item). Though similar to the method-factor model examined by McGee et al. (1989), the two-factor model examined in the current study goes further by positing a method factor composed of items from two scales, thus disentangling the potential effects of item wording from those attributable to item content. Finally, the fourth model posits the existence of two method factors: one factor defined by all the positively worded items, and a second defined by all the negatively worded items. This model is based on the plausible suggestion that all of the observed covariance between items is attributable to method variance (i.e., the direction of item wording).

A second issue arising from McGee et al.'s (1989) analysis is addressed in this study. Although their method-factor model provided the best fit to the data, this is to be expected because the fit of a model improves with the increase in the number of parameters estimated (Mulaik et al., 1989). A more pertinent question is whether a two-factor (a general role-stress and a method factor) or a three-factor (as in the present study) model provides a more parsimonious fit to the data (James, Mulaik, & Brett, 1982; Mulaik et al., 1989). We calculated the parsimonious fit index (PFI; James et al., 1982) for the data presented by McGee et al. (1989), and the results suggest that the two-factor conceptual model with 29 estimated parameters (for three samples, PFIs = .70, .70, and .72) was superior to models hypothesizing one conceptual factor (28 estimated parameters; PFIs = .58, .58, and .60) or a general and a method factor with 34 estimated parameters (PFIs = .68, .69, and .69). We also tested the parsimony of the four models in this study because it would be counterproductive to achieve an increased goodness of fit for the three-factor model at the expense of parsimony. Finally, Cook and Campbell (1979) advise against using data from a single organization, as this would limit any inferences about the external validity of the findings. Consequently, in the present study, we used data from two independent samples.

Method

Subjects

Seven hundred and sixty-seven employees of a hospital in Ontario, Canada comprised the first sample. Approximately 85% of the sample were women; 76% were full-time employees. The single largest occupational group was nursing (approximately 43% of the sample). The mean age of the sample was 36 years (range: 18–72), mean length of time with the organization was 8 years (range: 1–35), and educational attainment ranged between noncompletion of high school to postgraduate training (M = 15 years of formal education).

Data for the second sample were taken from the first wave of a four-wave study of work and family issues. Subjects were recruited randomly across Canada by a market research agency, which provided to us the names of 323 couples willing to participate. We mailed these couples questionnaires and a letter explaining the nature and purpose of the study. Two hundred and twenty-four men (69%) and 223 women (69%) returned usable questionnaires. The data used here are based on a subset of this sample, namely, those 363 respondents who were employed outside the home on a full-time (95%) or part-time (5%) basis. Their mean age was 38.02 years (range: 22–66), they had been with their employers for an average of 8.82 years (range: 0–37), and they had completed an average of 13.31 years of education (again ranging from noncompletion of high school through postgraduate training). Subjects' monthly incomes were considerably heterogeneous: (M = $4,110.29, range: $930–$18,500 [Canadian dollars]).

Questionnaires

All respondents completed the eight-item Role Conflict Scale and the six-item Role Ambiguity Scale, developed by Rizzo et al. (1970), and the three-item Role Overload Scale, developed by Beehr et al. (1976). The internal consistency of the conflict and ambiguity scales was satisfactory for both samples (hospital sample: role conflict = .82, role ambiguity = .80; random sample: role conflict = .70, role ambiguity = .80). The internal consistency of the overload scale was less satisfactory (hospital sample = .49, random sample = .69), possibly because of its multidimensionality (e.g., qualitative and quantitative role overload) or the small number of items or both.

Data Analysis

Following McGee et al. (1989), we computed the item variance–covariance matrix for all analyses. Within each sample, five models were estimated with maximum likelihood estimation as implemented in LISREL VI (Jöreskog & Sörbom, 1981). In addition to the four models described earlier, a null model specifying no common factors was estimated to provide a basis for comparison of the three substantive models of interest. Fit indices included those available through LISREL VI together with the normed fit index (NFI; Bentler & Bonett, 1980) and the PFI (James et al., 1982). The chi-square:degrees of freedom ratio was also calculated.

The difference in fit between nested models can be evaluated with a chi-square test in which the degrees of freedom are equal to the differences in the degrees of freedom of the chi-square values of the two models being contrasted (Long, 1983). For the current study, the null model was nested in all models of substantive interest, and the one-factor model was nested within both two-factor models and the three-factor model. That is, by constraining the interfactor correlations to equal 1 in the three-factor model and the two-factor model (positive and negative wording), the one-factor model can be reproduced. Similarly, the one-factor model can be reproduced from the other two-factor model (role stress and negative wording) by constraining all loadings on the second factor to equal 0. Chi-square tests were calculated to examine these differences. Neither two-factor model was nested within the three-factor model, and therefore no statistical comparison of the goodness of fit of these two models was appropriate.

Results

Fit indices for the null model and the three substantive models of interest are presented in Table 1. For both samples, the null model offered a poor fit to the data. The one-factor model represented a significant improvement in fit over the null model, according to the test for the difference between models: $\chi^2(17, N = 767) = 2{,}401.83$, $p \leq .001$, for the hospital sample;

Table 1
Fit Indices for Confirmatory Factor Analysis of the Four Models

Model	χ^2	*df*	χ^2/df	GFI	AGFI	RMSR	NFI	PFI
			Sample 1 (N = 767)					
Null	3,775.63	136	27.76	.45	.37	.26	—	—
One factor	1,373.80	119	11.55	.74	.66	.10	.64	.56
Two factors[a]	733.62	112	6.56	.89	.85	.08	.81	.67
Two factors[b]	549.16	118	4.65	.92	.90	.07	.86	.75
Three factors	400.48	116	3.45	.94	.92	.06	.89	.76
			Sample 2 (N = 363)					
Null	1,840.69	136	13.53	.49	.42	.24	—	—
One factor	858.91	119	7.22	.73	.66	.12	.53	.46
Two factors[a]	456.06	112	4.07	.87	.82	.09	.75	.62
Two factors[b]	567.68	118	4.81	.84	.80	.12	.69	.60
Three factors	413.46	116	3.56	.88	.84	.09	.78	.66

Note. GFI = goodness-of-fit index. AGFI = adjusted GFI. RMSR = root-mean-square residual; NFI = normed fit index; PFI = parsimonious fit index. The NFI and PFI were not applicable to the null model. [a] One general role-stress factor (all 17 items) and a second, orthogonal method factor (all 7 negatively worded items, regardless of content). [b] One factor representing the 7 negatively worded items and one factor representing the 10 positively worded items.

$\chi^2(17, N = 363) = 981.78$, $p \leq .001$, for the random sample. The two-factor model (role stress and negative wording) represented an improvement in fit over the one-factor model—$\chi^2(7, N = 767) = 640.18$, $p \leq .001$, for the hospital sample; $\chi^2(7, N = 363) = 402.85$, $p \leq .001$, for the random sample—as did the model hypothesizing a factor comprising all positively worded items and a factor comprising negatively worded items—$\chi^2(18, N = 767) = 824.64$, $p \leq .001$, for the hospital sample; $\chi^2(18, N = 363) = 291.23$, $p \leq .001$, for the random sample. Likewise, the three-factor model provided a significantly better fit to the data than the one-factor model: $\chi^2(3, N = 767) = 973.32$, $p \leq .001$, for the hospital sample; $\chi^2(3, N = 363) = 445.45$, $p \leq .001$, for the random sample.

All the fit indices indicate that the three-factor model provides the most appropriate solution. Although direct statistical comparisons are not appropriate between the two- and three-factor models, the three-factor model was associated with lower chi-square and chi-square:degree of freedom values and higher goodness-of-fit-index (GFI) and adjusted GFI (AGFI) values than any of the competing models. In the hospital sample, the value of the NFI for the three-factor model approached the recommended value of .90 (Bentler & Bonett, 1980). In the random sample, the NFI was larger for the three-factor model (NFI = .78) than for the one-factor model (NFI = .53) or either two-factor model (NFI = .75 for role stress and negative wording; NFI = .69 for positive and negative wording). The three-factor model also provided the most parsimonious solution in both samples, manifesting the highest PFI values (.76 for the hospital sample; .66 for the random sample) of the four models considered.

The model parameters of the three-factor model (standardized solution) are presented in Table 2. In both samples, parameter estimates were associated with a small standard error of estimate and, as a result, were statistically significant. The model parameters suggest that one conflict item ("I have to do things that should be done differently") was only marginally associated with the role conflict dimension in the random sample and was not significantly associated with the conflict dimension in the hospital sample.

Discussion

The results of this study offer little support for either of the hypotheses suggesting that Rizzo et al.'s (1970) role conflict and ambiguity scales suffer from a lack of construct validity. Specifically, little support emerged for a general role-stress dimension reflected in a single-factor model (Tracy & Johnson, 1981). Likewise, although operationalized in two different ways, little empirical support emerged for the hypothesis that the confound between item wording and item content (McGee et al., 1989) limits the interpretability of Rizzo et al.'s (1970) scales. Instead, the model that received the most empirical support is consistent with the conceptual definitions of role conflict, role ambiguity, and role overload.

These results supporting the construct validity of the role ambiguity and conflict scales can be considered robust for several reasons. First, support for the three-factor model was consistent across the various goodness-of-fit tests computed. Second, the three-factor model was the most parsimonious; based on 36 estimated parameters, the three-factor model provided a better fit than did the one-factor (34 parameters) or either of the two-factor (41 and 35 parameters) models. Third, although external validity is ultimately a matter of continual replication, inferences about the generalizability of these findings are justified because of the empirical support found in two large, independent samples for the three-factor model.

Support for the construct validity of the role conflict and ambiguity scales is of some conceptual importance. A considerable body of knowledge about role stressors is based on Rizzo et al.'s (1970) scales (e.g., Fisher & Gitelson, 1983; Jackson & Schuler, 1985). If the Rizzo et al. scales are compromised by a lack of construct validity, severe implications for the integrity

Table 2
Model Parameters for the Three-Factor Model: Standardized Solution

Item	Factor loading			h^2
	Ambiguity	Conflict	Overload	
I feel secure about how much authority I have.				
Hospital sample	.55**			.30
Random sample	.38**			.14
Clear, planned goals and objectives exist for my job.				
Hospital sample	.59**			.35
Random sample	.44**			.19
I know that I have divided my time properly.				
Hospital sample	.48**			.23
Random sample	.30**			.09
I know what my responsibilities are.				
Hospital sample	.78**			.60
Random sample	.77**			.59
I know exactly what is expected of me.				
Hospital sample	.85**			.72
Random sample	.85**			.72
Explanation is clear of what has to be done.				
Hospital sample	.74**			.55
Random sample	.64**			.41
I have to do things that should be done differently.				
Hospital sample		.04		.002
Random sample		.11*		.01
I receive an assignment without the manpower to complete it.				
Hospital sample		−.61**		.37
Random sample		−.71**		.51
I work with two or more groups who operate quite differently.				
Hospital sample		−.52**		.26
Random sample		−.54**		.29
I have to buck a rule or policy to carry out an assignment.				
Hospital sample		−.71**		.50
Random sample		−.70**		.49
I receive incompatible requests from two or more people.				
Hospital sample		−.76**		.58
Random sample		−.66**		.44
I do things that are apt to be accepted by one person and not accepted by others.				
Hospital sample		−.65**		.42
Random sample		−.59**		.35
I receive an assignment without adequate resources and materials to execute it.				
Hospital sample		−.74**		.55
Random sample		−.77**		.60
I work on unnecessary things.				
Hospital sample		−.52**		.27
Random sample		−.45**		.21
I am given enough time to do what is expected of me on my job.				
Hospital sample			.59**	.34
Random sample			.50**	.25
It often seems like I have too much work for one person to do.				
Hospital sample			.43**	.19
Random sample			.62**	.38
The performance standards on my job are too high.				
Hospital sample			.42**	.18
Random sample			.67**	.44

* $p < .05$. ** $p < .01$.

of this knowledge base would ensue. Because the present results suggest consistently that a methodological confound is not a plausible interpretation when item wording and item content are disentangled, the integrity of prior findings using Rizzo et al.'s (1970) scales is not threatened. However, a more direct test of the effect of negative item wording would be to rewrite items comprising the role conflict and ambiguity scales so that each scale comprises both positively and negatively worded items. This modification would allow for the examination of method and trait variance within a multitrait–multimethod paradigm (e.g., Widaman, 1985).

Support for the construct validity of the role conflict and

ambiguity scales does not preclude the necessity of developing better measures (e.g., McGee et al., 1989; Tracy & Johnson, 1981). At least one item in the Role Conflict Scale ("I have to do things that should be done differently") was only marginally associated with other items in the scale. Examination of the modification indices suggests that this item might be more appropriately assigned to the Role Ambiguity Scale. This reclassification is inconsistent with the hypothesis of a method factor (e.g., McGee et al., 1989) but may be consistent with the conceptual definition of role ambiguity. That is, although the item reflects the presence of role stress (rather than the absence of role stress as do the other ambiguity items), it may be interpreted as representing ambiguity in role demands.

A word of caution concerning the use of negatively worded items is appropriate. Because their use may introduce unintended covariance (Schmitt & Stults, 1985), some researchers have begun to question the utility of negatively worded items (e.g., Thacker, Fields, & Tetrick, 1989). Nonetheless, the present study suggests that covariance introduced by the direction of item wording does not necessarily result in a methodological confound that distorts conceptual interpretation. If negatively worded items are to be used, however, it would be wise to ensure during scale development that their inclusion does not present a methodological confound. Rather than having alternative interpretations to contend with some two decades after scale development (e.g., McGee et al., 1989; Tracy & Johnson, 1981), it would be preferable to ensure that constructs are not exclusively defined by negatively worded items during scale development. Instead, scales should have equal numbers of positively worded and negatively worded items.

We examined the potential effect of construct-irrelevant covariance on the interpretability of role ambiguity and role conflict, but other concerns regarding the content, convergent, and discriminant validity of these scales remain (King & King, 1990), which suggests a need for ongoing concern with measurement techniques. In the interim, however, the present findings suggest that the existing measures of role conflict and role ambiguity (Rizzo et al., 1970) exhibit construct validity and can be interpreted meaningfully. Calls for a moratorium on the use of the Role Conflict Scale and the Role Ambiguity Scale, therefore, are premature.

References

Beehr, T. A., Walsh, J. T., & Taber, T. D. (1976). Relationship of stress to individually and organizationally valued states: Higher order needs as a moderator. *Journal of Applied Psychology, 61,* 41–47.

Bentler, P. M., & Bonett, D. C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Cook, J. D., Hepworth, S. J., Wall, T. D., & Warr, P. B. (1981). *The experience of work: A compendium and review of 249 measures and their use.* San Diego, CA: Academic Press.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Fisher, C. D., & Gitelson, R. (1983). A meta-analysis of the correlates of role conflict and ambiguity. *Journal of Applied Psychology, 68,* 320–333.

Jackson, S. E., & Schuler, R. S. (1985). A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings. *Organizational Behavior and Human Decision Processes, 36,* 16–78.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models and data.* Beverly Hills: Sage.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables and least squares methods.* Mooresville, IN: Scientific Software.

Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. (1964). *Organizational stress: Studies in role conflict and ambiguity.* New York: Wiley.

King, L. A., & King, D. W. (1990). Role conflict and role ambiguity: A critical assessment of construct validity. *Psychological Bulletin, 107,* 48–64.

Long, J. S. (1983). *Confirmatory factor analysis: A preface to LISREL.* Beverly Hills: Sage.

McGee, G. W., Ferguson, C. E., & Steers, A. (1989). Role conflict and role ambiguity: Do the scales measure these two constructs? *Journal of Applied Psychology, 74,* 815–818.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennet, N., Lind, S., & Stillwell, C. D. (1989). Evaluation of goodness of fit indices for structural equation models. *Psychological Bulletin, 105,* 430–445.

Rizzo, J. R., House, R. J., & Lirtzman, S. I. (1970). Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly, 15,* 150–163.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61.

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement, 9,* 367–373.

Thacker, J. W., Fields, M. W., & Tetrick, L. (1989). The factor structure of union commitment: An application of confirmatory factor analysis. *Journal of Applied Psychology, 74,* 228–232.

Tracy, L., & Johnson, T. W. (1981). What do the role conflict and role ambiguity scales measure? *Journal of Applied Psychology, 66,* 464–469.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait–multimethod data. *Applied Psychological Measurement, 9,* 1–26.

Received February 1, 1990
Revision received April 9, 1990
Accepted May 31, 1990 ■