Human Resource Management Review 18 (2008) 223–232

Contents lists available at ScienceDirect

Human Resource Management Review

journal homepage: www.elsevier.com/locate/humres

Rater source factors represent important subcomponents of the criterion construct space, not rater bias

Charles E. Lance [a,*], Brian J. Hoffman [a], William A. Gentry [b], Lisa E. Baranik [a]

[a] The University of Georgia, United States
[b] Center for Creative Leadership, United States

ARTICLE INFO

Keywords:
Multisource performance ratings
Multitrait-multimethod
MTMM
Confirmatory factor analysis
Psychometrics

ABSTRACT

We contrast normative accuracy and ecological perspectives on applications of the multitrait–multimethod methodology to multisource performance ratings and review existing research that provides critical tests of these perspectives. Existing research supports the ecological perspective which proposes that the rater source effects that are typically found in analysis of multisource performance ratings do not represent mere halo biases but alternative, perhaps equally valid perspectives on ratee performance. We suggest that future research view multifaceted research designs in the broader context of a prototype multidimensional data relational system such as that proposed by Lance, Baranik, Lau, and Scharlau (Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E. A. (in press). If it's not trait it must be method: (Mis) application of the multitrait–multimethod design in organizational research. In C. E. Lance & R. J. Vandenberg (Eds.), Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences. Mahwah, NJ: Erlbaum).

Productivity and job performance are cornerstone constructs in human resource management (Borman, 1991; Pritchard, 1992). For example, job performance measurement is an integral component in test validation research (Campbell, 1990; McDonald, 1999), training needs assessment and training program evaluation (Arvey & Cole, 1989; Campbell, 1988; Ostroff & Ford, 1989), promotion and succession planning (Burack & Mathys, 1987), salary administration (Cascio, 1989; Hammer, 1988), recruitment, selection and placement (Burke & Pearlman, 1988; Schneider & Schmitt, 1992), and developmental feedback toward performance maintenance and improvement (Smither, London, & Reilly, 2005).

The last two decades have seen a dramatic increase in popularity of a particular performance measurement and feedback approach that is known as multisource performance rating (MSPR; Conway & Huffcutt, 1997) or 360-degree assessment and feedback (London & Tornow, 1998). In a typical application of a MSPR program managers are rated by their supervisors, peers, and subordinates (and perhaps also by themselves and their clients), ratings are aggregated within each source where there are multiple raters, and then developmental feedback with respect to the aggregated ratings on relevant performance dimensions is given to ratees for performance review and planning purposes (Church & Bracken, 1997; London & Smither, 1995; Smither et al., 2005). From an applied perspective, MSPRs are thought to be valuable in part because ratings from different sources provide complementary views of the ratee's performance from different organizational perspectives (Borman, 1997). However, from a traditional psychometric perspective, research on MSPRs has consistently produced what has been interpreted as representing a pattern of troubling findings: despite the fact that ratings *within* sources have some demonstrated convergence, almost invariably there is low to moderate convergence (at best) in ratings *across* sources (Conway, 1996; Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; LeBreton, Burgess, Kaiser, & James, 2003). This relative lack of convergence across sources in MSPRs has been viewed as reflecting (a)

* Corresponding address. Department of Psychology, The University of Georgia, Athens, GA 0602-3013, United States. Tel.: +1 706 542 3053; fax: +1 706 542 3275.
E-mail address: clance@uga.edu (C.E. Lance).

1053-4822/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.hrmr.2008.03.002

valued complementary perspectives on ratee performance by 360-degree feedback practitioners and, alternately, (b) the influences of pervasive unwanted, contaminating, rater source *bias* by those who have researched MSPRs from a traditional psychometric perspective. The purpose of this paper is to attempt a resolution of these conflicting interpretations of the same evidential data base.

We argue in this article that MSPR researchers may well have been misled into thinking that rater source effects represent rater *biases* through their cavalier adaptation of the multitrait–multimethod (MTMM) methodology to study the latent structure of MSPRs. Applications of the MTMM methodology to the study of MSPRs have routinely assumed that (a) the rating dimensions under study represent the *Traits* in the MTMM design, and (b) since rater sources do not represent the Traits of interest, they must therefore represent the *Methods* in the MTMM design. Combined with these assumptions and typical empirical findings that (a) correlations between the same performance dimensions as rated by different sources (corresponding to monotrait-heteromethod correlations in an MTMM matrix) are relatively low, and (b) correlations between different performance dimensions as rated by the same source (corresponding to heterotrait-monomethod correlations in an MTMM matrix) are relatively high, MSPRs have been assumed to demonstrate weak convergent validity (at best), and strong and pervasive measurement method effects, from a traditional psychometric perspective. In the remainder of this article we present background on the MTMM methodology, discuss two competing theories on the nature of rater source effects in MSPRs, review literature that has pitted these competing theories against one another empirically, and discuss implications of these findings both from scientific and applied perspectives.

1. Background

The MTMM matrix was introduced by D. T. Campbell and Fiske (1959) as an innovative approach to the study of convergent and discriminant validity of psychological measures. This article is now one of the most often cited in psychology (Fiske & Campbell, 1992) — as of 31 October 2007 it had been cited 4450 times in the Web of Science alone in a wide variety of disciplines including the social and physical sciences, law, medicine and education (see Lance et al., in press). One of the reasons for the widespread adoption of the MTMM methodology is that the provision of convergent and discriminant validity evidence is widely regarded as a cornerstone for the establishment of measures' construct validity (Benson, 1998; Messick, 1995).

D. T. Campbell and Fiske's (1959) criteria for convergent and discriminant validity and the presence of method effects are now widely recognized as being rather subjective (Widaman, 1985). As a result, a number of more objective, quantitative approaches to the analysis of MTMM matrices have developed over the years, including analysis of variance (e.g., Boruch, Larkin, Wolins, & McKinney, 1970; Kavanagh, MacKinney & Wolins, 1971), path analysis (Avison, 1978; Schmitt, 1978) multiple regression (e.g., Lehmann, 1988), and exploratory factor analysis (Golding & Seidman, 1974; Wothke, 1995). Today however, the most popular analytic model of choice is some form of a confirmatory factor analysis (CFA) model. A number of such models have been proposed including a family of additive models that specify latent Trait and Method factors' effects on observed measures (e.g., Widaman, 1985), models that include latent Trait factors but which model method effects as covariances among uniquenesses of measures of traits using the same measurement method (e.g., Marsh, 1989), hierarchical CFA models (e.g., Marsh & Hocevar, 1988), models that specify interactive Trait x Method effects (Browne, 1984; Campbell & O'Connell, 1967), in addition to others (e.g., Eid, 2000; Kenny & Kashy, 1992; Lance, Woehr, & Meade, 2007). The first two of these, sometimes referred to as the correlated trait-correlated method (CTCM) and the correlated trait-correlated uniqueness (CTCU) models, are the most widely accepted and implemented models. The CTCM model can suffer convergence and admissibility problems (Brannick & Spector, 1990; Kenny & Kashy, 1992) that the CTCU model often avoids, but Conway, Lievens, Scullen, and Lance (2004) and Lance, Noble and Scullen (2002) showed that the CTCU model suffers from a number of other serious conceptual and analytic problems. Consequently we only consider the general CTCM model (and several of its special cases) here.

2. An Illustration of the CTCM Model for MSPRs

As an illustration of applying the MTMM framework to MSPRs, we obtained data from a sample of 22,420 managers who were rated on the Center for Creative Leadership's Benchmarks®[1] (Center for Creative Leadership, 2004; Lombardo & McCauley, 1994; McCauley, Lombardo, & Usher, 1989) multisource feedback instrument by their supervisor, peers, subordinates and themselves. Benchmarks® consists of 16 separate measures, but in the interest of parsimony we combined these into three broad performance dimensions (Meeting Job Responsibilities, Respecting Self and Others, and Leading People) that map onto a generalized taxonomy of managerial performance developed by Borman and Brush (1993) and that has previously been used with the Benchmarks® instrument (see Fleenor, McCauley, & Brutus, 1996). Also, although common practice is to aggregate ratings within sources when presenting feedback, for illustrative purposes we randomly selected (a) one peer and one subordinate, and (b) a subset of 520 ratees for analysis. Thus the subsample reported here consists of 520 managers who provided self-ratings and who were rated by their supervisor and one peer and subordinate each.

Correlations among Benchmarks® ratings are shown in Table 1. A preliminary (and subjective) analysis of these ratings indicates some support for convergent validity: the average different source-same dimension correlation (mean r=.18) is somewhat larger than the mean different source-different dimension correlation (mean r=.13, $t(52)$=4.52, $p<.01$), although this difference is not large in an absolute sense. On the other hand, there is evidence of strong rating source effects: the mean different dimension-same source correlation (mean r=.81) is substantially higher than either the average same dimension-different source correlation (mean r=.18, $t(28)$=33.96, $p<.01$) or the average different dimension-different source correlation (mean r=.13, $t(46)$=39.74, $p<.01$).

[1] Benchmarks® is a registered trademark of the Center for Creative Leadership.