

**CONSTRUCT VALIDATION STANDARDS
AND THE TEAM DESCRIPTIVE INDEX: REPLY TO DOWLING**

Stephanie M. Lee, *Baylor University*

Joel Koopman, *Texas A&M University*

John R. Hollenbeck, *Michigan State University*

Linda C. Wang, *City University of Hong Kong*

Klodiana Lanaj, *University of Florida*

Abstract

For the last 60 years, scale development efforts in Psychology and Management have closely followed psychometric theory and procedures outlined by Cronbach and Meehl (1955). While many have improved and positively added to psychometric practices (e.g., Bentler, 1990; Campbell & Fiske, 1959; Jöreskog, 1969; Nunnally, 1978; Reckase, 2009; Schwab, 1980), there have been attempts to develop alternative methods. One such approach is the C-OAR-SE procedure (Rossiter, 2002), which was recently used to evaluate the Team Descriptive Index (TDI; Lee et al., 2015) in “Commentary on the Team Descriptive Index” (Dowling, in press). In this response to the commentary, we offer a peer-review of the C-OAR-SE procedure and arguments for why we stand by the psychometric theory underlying the development of the TDI. It is our hope that our response will have value to future scholars who have utilized traditional psychometric methods to construct validate measures and face critiques on their methodology as we witnessed in “Commentary on the Team Descriptive Index.”

Keywords: Team composition, construct validity, measurement, validation

Construct Validation Standards and the Team Descriptive Index: Reply to Dowling

The psychometric theory and practices that underlie the development of the Team Descriptive Index (TDI) have been embraced in Psychology and Management research for over 60 years. It is safe to say that almost all scale development efforts follow the procedures outlined by Cronbach and Meehl (1955), and later improved by others (Bentler, 1990; Campbell & Fiske, 1959; Jöreskog, 1969; Nunnally, 1978; Reckase, 2009; Schwab, 1980). These procedures are standard content in virtually all research methods courses. The clear and current consensus is that construct development needs to rely on a joint theoretical and empirical process where multiple forms of evidence from different sources, collected at different times using different methods, all provide important pieces of a single puzzle (Hinkin, 1995; Boateng, Neilands, Frongillo, Melgar-Quinonez, & Young, 2018).

This process typically involves conceptualizing the construct in its nomological net, followed by a set of traditional empirical procedures. These procedures would include:

- (a) various forms of factor analyses of potential items to appropriately identify construct dimensions;
- (b) testing of internal consistency through an examination of test-item correlations (e.g., Coefficient alpha);
- (c) collection of evidence regarding test-retest reliability if the construct is supposedly stable;
- (d) assessment of inter-rater reliability evidence if the construct is socially constructed;
- (e) procurement of content validity evidence from expert raters attesting to item-construct overlap;

(f) convergent validity as it relates to other constructs that should be similar to the focal construct;

(g) discriminant validity with respect to constructs that differ from the focal construct, a

(h) criterion-related validity evidence as it pertains to predicting important future outcomes.

The current consensus notwithstanding, there have been attempts to develop alternative methods for scale development—attempts that have largely failed to gain traction. The C-OAR-SE procedure reflects one such effort (Rossiter, 2002). C-OAR-SE originated in the Marketing literature as a reaction to what Rossiter saw as an overreliance on statistical evidence in scale development. Rather than being rooted in traditional psychometrics, C-OAR-SE rejects psychometrics; indeed, Rossiter (2002: 308) notes C-OAR-SE is “grounded in rationalism rather than empiricism.” For scale development, C-OAR-SE thus eschews multiple forms of empirical validity evidence (such empirics are described as “irrelevant,” “immaterial,” and “misleading”; Rossiter, 2011: 1565) in favor of expert evaluations of content validity alone. While content validity is indeed critical, Rossiter’s peers in the field of Marketing have argued C-OAR-SE is insufficient “to guarantee a measurement instrument’s validity” (Salzberger, Diamantopoulos, & Sarstedt, 2016: 1945).

While the C-OAR-SE procedure remains largely unused in the very field in which it was developed, the present commentary by Dowling (in press) seeks to export this procedure to the Management literature via the *Academy of Management Discoveries (AMD)* Commentary Section. Whereas the article in question — Lee, Koopman, Hollenbeck, Wang, and Lanaj (2015) — underwent three rounds of peer review at *AMD* (and was Runner-Up for Best Paper in 2015), the *AMD* Commentary Section is not similarly peer-reviewed. Thus, we offer our rebuttal in an

effort to provide the necessary context for and peer review of the Dowling's (in press) critique of the TDI.

We begin by briefly describing the C-OAR-SE procedure, as we suspect it is unfamiliar to scholars outside of Marketing. C-OAR-SE calls for a construct to be defined in terms of (1) an object, (2) an attribute, and (3) a rater. As an example, Rossiter (2002: 309) uses service quality. Service quality is only an attribute of a construct because “one cannot conceive of service quality in the abstract.” Instead, it needs an object (e.g., “service quality in America” or “IBM’s service quality”). Finally, constructs need a rater because ratings “cannot be divorced from the perceiver...the rater entity is part of the construct” (Rossiter, 2002: 318). Note this implicitly views constructs as idiosyncratic and narrow, rather than broad and generalizable. Following this, items are written. Of note, C-OAR-SE advocates the use of single-item measures for concrete constructs, or for each component of abstract constructs (concrete vs. abstract being a determination about the object made by expert judges). C-OAR-SE recommends using degree, probability, or frequency response formats, with a “don’t know” option and equally-spaced intervals between response points. The rules for scale enumeration are complex (Table 3; Rossiter, 2002: 324) and vary based on the classification of both the object and attribute.

Rossiter (2002: 331) noted C-OAR-SE is a “theoretical and procedural solution” for scale development. He also noted that this procedure would likely not be received well. Based on the past 17 years of subsequent scholarship, he was half right. C-OAR-SE has indeed not been received well (e.g., Diamantopoulos, 2005; Finn & Kayande, 2005; Salzberger et al., 2016; Sarstedt, Diamantopoulos, Salzberger, 2016a). However, not only is C-OAR-SE rarely applied in practice, but most of the citations to Rossiter’s seminal article have been used to “justify the use of single items” (Sarstedt, Diamantopoulos, Salzberger, 2016b: 3200). Tabling for the moment

the appropriateness of C-OAR-SE's advocacy for properly developed single-item measures, as Sarstedt et al. (2016b) note, most manuscripts are not actually implementing C-OAR-SE. Thus, C-OAR-SE's biggest contribution thus far has been enabling wishful thinking by authors seeking to justify poorly operationalized constructs. Single-item measures are widely known to be problematic and conceptually deficient. As many have both argued and repeatedly demonstrated, single-item measures are unreliable, produce numerous Type II Error, and fail to promote generalization (e.g., Nunally, 1978). We recommend *AMD* readers not follow this lead.

C-OAR-SE's emphasis on specifying the object, attribute, and rater for a scale forces researchers to define constructs so narrowly as to undermine an instrument's utility (e.g., making comparisons between studies). Yet there is something even more alarming here. C-OAR-SE relies upon expert raters to evaluate the concreteness vs. abstractness of a construct. As Sarstedt et al. (2016b) note, extant theory on the "validity" of constructs developed with the C-OAR-SE procedure (e.g., Bergkvist & Rossiter, 2007, 2009; Rossiter, 2002, 2011, 2012) uses the same two expert raters: Bergkvist and Rossiter. The shunning of empirical evidence, coupled with the reliance upon the presumed expert status of a scale developer, charts a troubling course for the future of measure development and construct validation. As classical test theory recognizes, each type of validity evidence provides a single piece of the construct validation puzzle. Developing scales with multiple items and evidentiary bases ensure the development of scales that actually measure what they are intended to measure.

As it applies to the TDI, applying the C-OAR-SE procedure to measuring team types would actually undo the problem that Hollenbeck, Beersma, and Schouten (2012) initially set out to remedy. As these authors note, the teams literature was afflicted with a "plethora of alternative team taxonomies and no consensus regarding how to describe or classify teams" (Hollenbeck et

al., 2012: 82). The theory from Hollenbeck and colleagues, combined with the scale developed by Lee et al. (2015), alleviated this by focusing on three underlying, generalizable dimensions through which all teams (regardless of occupation, purpose, industry, specialization, or tenure) could be classified. Taking Dowling's critique at face value and applying C-OAR-SE to the TDI would undo this, returning the teams literature to the days of ad-hoc categorization schemes.

In conclusion, we stand behind the psychometric theory and practices that informed our scale development efforts on the TDI. We also note that, although Dowling chose to single us out for criticism, none of the criticisms he levied are, in fact, unique to the TDI. Indeed, these criticisms are equally applicable to scales such as the Multifactor Leadership Questionnaire (Bass & Avolio, 1988), the Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988), or the recently developed Team Processes Survey Measure (Mathieu, Luciano, D'Innocenzo, Klock, and LePine, 2019). While the TDI can certainly be improved, we do not feel that the Dowling's critique, or the C-OAR-SE procedure for that matter, offer any practical value. Not only is the C-OAR-SE procedure woefully deficient regarding scale development, it is misguided in its endorsement of single-item measures. Scholars that adopt the single-item norm are not likely to find that this is acceptable in peer-reviewed journals. This is a conclusion already well-recognized in Marketing literature—and now, perhaps—the Management literature as well.

References

- Bass, B. M. & Avolio, B. J. 1989. *Manual for the Multifactor Leadership Questionnaire*. Palo Alto, CA: Consulting Psychologists Press.
- Bentler, P. M. 1990. Comparative fit indexes in structural models. *Psychological Bulletin*, 107: 238–246.
- Bergkvist, L., & Rossiter, J. R. 2007. The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, 44, 175–184.
- Bergkvist, L., & Rossiter, J. R. 2009. Tailor-made single-item measures of doubly concrete constructs. *International Journal of Advertising*, 28, 607–621.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. 2018. Best practices for developing and validating scales for health, social, and behavioral research: A Primer. *Frontiers of Public Health*, 6: 149–166.
- Campbell, D. T., & Fiske, D. W. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56: 81–105.
- Cronbach, L. J., & Meehl, P. E. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52: 281–302.
- Diamantopoulos, A. 2005. The C-OAR-SE procedure for scale development in marketing: A comment. *International Journal of Research in Marketing*, 22: 1–9.
- Dowling, G. (in press). Commentary on the team descriptive index. *Academy of Management Discoveries*. DOI: 10.5465/amd.2013.0001.
- Finn, A., & Kayande, U. 2005. How fine is C-OAR-SE? A generalizability theory perspective on Rossiter's procedure. *International Journal of Research in Marketing*, 22: 11–21.
- Hinkin T.R. (1995). A review of scale development practices in the study of organizations

- Journal of Management*, 21: 967–988.
- Hollenbeck, J. R., Beersma, B., & Schouten, M. E. 2012. Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. *Academy of Management Review*, 37: 82–106.
- Jöreskog, K. G. 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34: 183–202.
- Lee, S. M., Koopman, J., Hollenbeck, J. R., Wang, L. C. & Lanaj, K. 2015. The team descriptive index (TDI): A multidimensional scaling approach for team description. *Academy of Management Discoveries*, 1: 89-114.
- Mathieu, J. E., Luciano, M. M., D’Innocenzo, L., Klock, E. A., & LePine, J. E. 2019. The development and construct validity of a team processes survey measure. *Organizational Research Methods*. DOI: 10.1177-1094428119840801.
- Nunnally, J. C. 1978. *Psychometric Theory*. New York, NY: McGraw-Hill.
- Reckase, M. D. 2009. Multidimensional item response theory models. In *Multidimensional Item Response Theory*: 79–112. Springer, NY: Statistics for Social and Behavioral Sciences.
- Rossiter, J. R. 2002. The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19: 305–335.
- Rossiter, J. R. 2011. Marketing measurement revolution: The C-OAR-SE method and why it must replace psychometrics. *European Journal of Marketing*, 45, 1561–1588.
- Rossiter, J. R. 2012. A new C-OAR-SE-based content-valid and predictively valid measure that distinguishes brand love from brand liking. *Marketing Letters*, 23, 905–916.
- Salzberger, T., Diamantopoulos, A., & Sarstedt, M. 2016. Measurement in the social sciences: Where C-OAR-SE delivers and where it does not. *European Journal of Marketing*, 59:

1942–1952.

Sarstedt, M., Diamantopoulos, A., Salzberger, T., & Baumgartner, P. 2016a. Selecting single items to measure doubly concrete constructs: A cautionary tale. *Journal of Business Research*, 69, 3159–3167.

Sarstedt, M., Diamantopoulos, A., & Salzberger T. 2016b. Should we use single items? Better not. *Journal of Business Research*, 69: 3199–3203.

Schwab, D. P. 1980. Construct validity in organizational behavior. In B. M. Staw and L. L. Cummings (Eds.). *Research in Organizational Behavior*, 2: 3–43. Greenwich, CT: JAI Press.

Watson, D., Clark, L. A., & Tellegen, A. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54: 1063–1070.