



Contents lists available at ScienceDirect

## Journal of Business Research

Selecting single items to measure doubly concrete constructs: A cautionary tale<sup>☆</sup>Marko Sarstedt<sup>a,b,\*</sup>, Adamantios Diamantopoulos<sup>c</sup>, Thomas Salzberger<sup>d</sup>, Petra Baumgartner<sup>e</sup><sup>a</sup> Otto-von-Guericke-University Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany<sup>b</sup> University of Newcastle, Australia<sup>c</sup> University of Vienna, Oskar-Morgenstern-Platz 1, 1090, Vienna, Austria<sup>d</sup> Vienna University of Economics and Business, Augasse 2-6, 1090, Vienna, Austria<sup>e</sup> Home Shopping Europe GmbH, Münchener Str. 101h, 85737 Ismaning, Germany

## ARTICLE INFO

## Article history:

Received 1 March 2015

Received in revised form 1 September 2015

Accepted 1 October 2015

Available online xxxxx

## Keywords:

Single items

Multi-item scales

Measurement theory

Item response theory

C-OAR-SE

## ABSTRACT

Single-item measures have recently become more en vogue due to studies arguing in favor of their psychometric properties vis-à-vis multi-item scales. However, their effective use requires (1) expert raters to designate the focal construct as being doubly concrete and (2) researchers to find a good single item to represent the construct. This study examines whether expert raters identify the doubly concrete nature of constructs that prior research presents as exemplary in this respect. Furthermore, the study compares the efficacy of a broad range of selection mechanisms based on expert judgment and statistical criteria for identifying the best item in a scale. The results show that expert raters do not share the commonly held belief that researchers can validly measure constructs such as attitude toward the ad, or brand, with single items. Further analyses show that neither rater assessments nor statistical criteria prove valuable regarding identifying an appropriate single item from a set of candidate items.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

The extent to which successive generations of researchers and practitioners raise an issue is a subtle indicator of its importance. The benefits and limitations of single-item (SI) versus multi-item (MI) measures is one such issue that scholars have heatedly debated across a variety of disciplines as diverse as health care (e.g., Dolbier, Webster, McCalister, Mallon, & Steinhardt, 2005), sports management (Kwon & Trail, 2005), organizational psychology (e.g., Nagy, 2002), and marketing (e.g., Bergkvist, 2015, Bergkvist & Rossiter, 2007, 2009, Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012, Kamakura, 2015). Bergkvist and Rossiter (2007, 2009), who question the universal usefulness of MI measurement on theoretical and empirical grounds, produced two of the most prominent studies in this stream of research. Drawing on Rossiter's (2002, 2011) C-OAR-SE procedure, the authors argue that if the object of the construct (e.g., a brand, or an ad) is concrete and singular, and if its attribute (e.g., an attitude, or a perception) is also concrete, using an SI to

operationalize this construct is conceptually legitimate. In support of this view, Bergkvist and Rossiter (2007, 2009) report empirical findings showing that SI and MI measures are similar in terms of predictive validity and conclude that “carefully crafted single-item measures—of doubly concrete constructs—are at least as valid as multi-item measures of the same constructs, and that the use of multiple items to measure them is unnecessary” (Bergkvist & Rossiter, 2009, p. 618). Obviously, their recommendation is very attractive for researchers because the authors offer a theoretical and empirical justification for a simpler measurement, thereby producing substantial savings in data gathering efforts and costs. Unsurprisingly, authors increasingly justify their use of SIs by citing the article of Bergkvist and Rossiter (2007) (Kamakura, 2015), which has attracted over 460 SSCI citations as of January 2016.

At the same time, however, follow-up studies cast a different light on Bergkvist and Rossiter's (2007, 2009) work. For example, Diamantopoulos et al. (2012) systematically examine the influence of different measurement and data characteristics, such as the inter-item correlations, the number of items, and the sample size, on the predictive validity of SI versus MI measures. The authors conclude that “opting for SI measures in most empirical settings is a risky decision as the set of circumstances that would favor their use is unlikely to be frequently encountered in practice” (Diamantopoulos et al., 2012, p. 446). Specifically, researchers should consider SI measures (rather than an MI scale) only when (1) working with small sample sizes (i.e.,  $N < 50$ ), (2) the expected effect sizes are small, (3) the items of the originating MI scale are highly homogeneous (i.e., Cronbach's

<sup>☆</sup> A prior version of this paper was presented at the 2012 Academy of Marketing Science Annual Conference, New Orleans, USA. The authors thank Doreen Neubert and Kati Barth for their support with the analysis.

\* Corresponding author at: Otto-von-Guericke-University Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany. Tel.: +49 391 6718625.

E-mail addresses: marko.sarstedt@ovgu.de (M. Sarstedt), adamantios.diamantopoulos@univie.ac.at (A. Diamantopoulos), thomas.salzberger@wu.ac.at (T. Salzberger), Baumgartner.petra@gmx.de (P. Baumgartner).

alpha > 0.90), and (4) the items are semantically redundant. Similarly, Kamakura (2015) argues that Bergkvist and Rossiter (2007) confounded their results with spurious correlations due to common method biases and replicates their study taking measurement error into account. The results show that, once corrected for measurement errors, MI scales consistently outperform their SI equivalents.

Despite these discouraging results, researchers may still decide to opt for SI measures in the light of their manifold practical advantages (e.g., Drolet & Morrison, 2001, Fuchs & Diamantopoulos, 2009, Wanous, Reichers, & Hudy, 1997). In this case, however, two fundamental issues remain.

First, according to Bergkvist and Rossiter (2007, 2009), for a phenomenon to be amenable to SI measurement, a group of expert judges needs to identify the object and attribute that researchers are evaluating as being concrete. Researchers can consider an object as being concrete when virtually all the raters (e.g., respondents in a survey) “know what the object is and that, for them, there is only one object” (Rossiter, 2002, p. 311). Similarly, “a concrete attribute has virtually unanimous agreement by raters as to what it is, and they clearly understand that there is only one, or holistically one, characteristic being referred to when the attribute is posed, as in a questionnaire or interview, in the context of the to be rated object” (Rossiter, 2002, p. 313). To summarize, in order to justifiably use SI measures, a set of expert judges should agree that the construct under consideration is, in line with Rossiter’s (2002, 2011) C-OAR-SE procedure, indeed doubly concrete. However, to date, no study has evaluated whether expert judges share Bergkvist and Rossiter’s (2007, 2009) notion of the doubly concrete nature of such constructs as attitude toward the ad, or attitude toward the brand. The authors merely state that “according to expert judgment based on the C-OAR-SE procedure,  $A_{Ad}$  (or  $L_{Ad}$ ) and  $A_{Brand}$  [i.e., attitude toward the ad, likeability of the ad and attitude toward the brand] are two such constructs [i.e., doubly concrete constructs]” (Bergkvist & Rossiter, 2007, p. 176) without, however, providing any information on the number and qualification of the expert judges involved, or the procedure(s) used to elicit expert agreement. As noted elsewhere, “JR [John Rossiter] ignores his own advice” (Diamantopoulos, Sigauw, & Cadogan, 2008, p. 390).

Second, once researchers opt for SI measurement, they still need to decide on a particular item to operationalize the construct. Unfortunately, prior research has not reached consensus on this issue. For example, Rossiter (2002, p. 310) recommends using expert raters—that is, “[a] small group of judges with expertise regarding the construct”—to select the SI. Other researchers (e.g., Netemeyer, Bearden, & Sharma, 2003) recommend the use of actual respondents (i.e., naive raters) for scale reduction purposes who—unlike expert raters—offer an intuitive judgment without domain-specific background. Yet another approach based on judgment involves the creation of a free-standing, global SI on the grounds of content validity considerations (e.g., Bergkvist & Rossiter, 2009, Wanous et al., 1997). Finally, authors suggest using statistical criteria, such as indicator loadings (e.g., Loo, 2002, Sarstedt & Wilczynski, 2009). Methodological studies that empirically compare the validity of SI and MI measures considered several of these options (see Table 1). Some authors choose the item with the highest loading (e.g., Loo, 2002), others rely on expert raters (e.g., Sarstedt & Wilczynski, 2009) or create a global SI (e.g., Bergkvist & Rossiter, 2009), and still others state no rationale at all (e.g., Bergkvist & Rossiter, 2007). However, none of these authors offers a compelling argument as to why they favor a particular approach (or why they reject alternative approaches). Most importantly, no study has yet compared the efficacy of the different approaches to SI selection.

Against this background, the purpose of this study is twofold. First, the study assesses whether the key constructs that Bergkvist and Rossiter (2007, 2009) examine in their studies as prototypical of SI measurement can be considered doubly concrete. Following the guidelines in Rossiter’s (2002, 2011) C-OAR-SE procedure, the analysis draws on expert judgment to ascertain whether, or not, the focal constructs

have “a simple, clear object (e.g., an ad or a brand) and a single and single-meaning attribute (e.g., liking)” (Bergkvist, 2015).

Second, the study compares the efficacy of different SI selection approaches in finding the best SI; that is, the one with the highest predictive validity. While other types of validity (particularly discriminant validity) are also relevant when developing/selecting a measure, the current focus is on predictive validity because this criterion is particularly important for decision-making purposes (e.g., Aaker, Kumar, & Day, 2010; Crocker & Algina, 2008) and authors used it widely in prior empirical research comparing SI and MI measures (e.g., Bergkvist & Rossiter, 2007, 2009; Diamantopoulos et al., 2012; see also Table 1). Note that the item with the highest predictive validity does not necessarily match—let alone outperform—the predictive validity of an MI scale (Diamantopoulos et al., 2012). Using an SI is, more often than not, likely to be a second best solution, which, however, may be tolerable in light of the practical benefits of SIs (Fuchs & Diamantopoulos, 2009). Thus, the present study does not aim at making a case for SIs and acknowledges that the development of new measurement instruments arguably benefits from a MI operationalization (e.g., DeVellis, 2003, Netemeyer et al., 2003). SIs are also not suited for testing complex theories involving causal relationships and structural equation modeling (Hair, Sarstedt, Ringle, & Mena, 2012). In short, using a SI measure implicitly assumes that time- and cost-saving considerations take precedence over a more comprehensive, and thus generally more precise, measurement. The reality, however, is that authors do use SI measures widely in empirical research and they are therefore worthy of attention.

An evidence-based decision as to which SI a researcher should use requires an empirical investigation. However, evidently, comparing the performance of alternative SI measures relative to each other and against an MI measure runs contrary to the very idea of limiting data collection to just one item for the construct of interest. In any given study, administering an MI first and then selecting an SI would be nonsensical. A more reasonable approach is to disentangle SI selection and SI application in a two-stage procedure. In the first stage, an SI selection study investigates the performance of alternative SIs and, thus, informs subsequent application studies that administer a specific SI in substantive research endeavors. An SI selection study corresponds to a scale development project, which is also much more comprehensive than a scale’s subsequent application in practice. However, published evidence related to existing instruments may allow for conclusions in terms of potentially suitable SIs, provided that SI selection studies converge on specific SI selection criteria that are generalizable and included in typical scale development reports.

## 2. Methods and analysis

### 2.1. Overview

The study uses data from consumers’ responses to advertisements in order to assess the predictive validity of attitude toward the ad ( $A_{Ad}$ ) on attitude toward the brand ( $A_{Brand}$ ), as well as  $A_{Brand}$  on brand purchase intention ( $PI_{Brand}$ ). According to Rossiter’s (2002, 2011) C-OAR-SE procedure,  $A_{Ad}$ ,  $A_{Brand}$ , and  $PI_{Brand}$  are all doubly concrete constructs and should thus be measured with SIs (Bergkvist & Rossiter, 2007, 2009). Similarly, Bergkvist (2015) explicitly states that “examples of doubly concrete constructs include attitude toward the ad ( $A_{Ad}$ ), brand attitude ( $A_{Brand}$ ), and brand purchase intention ( $PI_{Brand}$ ).”

The analysis first centers on the question whether expert raters perceive the constructs under consideration as doubly concrete and, therefore, are amenable to SI measurement. Next, the analysis examines whether different SI selection approaches converge to the same item. The approaches include (1) judgment-based methods using expert raters, naive raters, and global SIs; (2) criteria based on classical test theory; and on (3) item response theory. Finally, the analysis evaluates whether the SIs that the selection approaches above identified are the ones with the highest predictive validity (i.e., the “best” items). Clearly,

Download English Version:

<https://daneshyari.com/en/article/10492630>

Download Persian Version:

<https://daneshyari.com/article/10492630>

[Daneshyari.com](https://daneshyari.com)