Christoph Fuchs/Adamantios Diamantopoulos

Using single-item measures for construct measurement in management research


Christoph Fuchs


Adamantios Diamantopoulos

Conceptual issues and application guidelines

Schlüsselbegriffe

Messtheorie; Reliabilität; Single-Item Skalen; Validität

Keywords

Measurement; reliability; single-item measures; validity

Zusammenfassung

Die Verwendung von Single-Item Messinstrumenten ist Gegenstand vermehrter Diskussionen in der aktuellen betriebswirtschaftlichen Forschung. Das Ziel dieses Beitrages liegt in der Entwicklung von konkreten Richtlinien, welche die Evaluierung des Einsatzausmaßes von Single-Item Messinstrumenten für die Operationalisierung von Konstrukten ermöglichen. Der Beitrag beginnt mit einer konzeptionellen Betrachtung von Single-Item Messinstrumenten und darauffolgend wird ein integratives Rahmenwerk entwickelt, mit Hilfe dessen die potentielle Akzeptanz von Single-Item Messinstrumenten anhand mehrerer relevanter Kriterien umfassend beurteilt werden kann.

Abstract

The use of single-item measures in management research has been subject to heavy debate in recent literature. This paper provides researchers with concrete guidelines on how to assess the extent to which a single-item measure can be legitimately used to operationalize the focal construct. We first present a conceptual perspective on single-item measures and follow this by an integrative framework within which the potential acceptability of single-item measures can be comprehensively evaluated on multiple relevant criteria.

Autoren

Christoph Fuchs, Assistant Professor, Department of Marketing and Statistics, Aarhus School of Business, University of Aarhus, Haslegårdsvej 10, Dk-8210 Aarhus, Denmark, Phone: +45-8948/6494, Fax: +45-8615/3988, E-Mail: chfu@asb.dk; Adamantios Diamantopoulos, Professor of International Marketing, Department of International Marketing, University of Vienna, Bruenner Strasse 72, A-1210 Vienna, Austria, Phone: +43-01/4277-38031, Fax: +43-1/4277-38034, E-Mail: adamantios.diamantopoulos@univie.ac.at.

1. Introduction

Surveys are the primary data collection method in management research (Bryman and Bell, 2003). The typical survey in this field seeks to generate data for testing relationships among several constructs; thus, a single construct is rarely covered in isolation. In this context, besides constructs being conceptualized as dependent and independent variables, most research designs also contain control and/or moderator variables, each of them usually measured with multi-item scales (e.g., Krosnick, 1999). This requires very lengthy survey instruments, which are likely to overload respondents (e.g., Wanous, Reichers and Hudy, 1997), lead to a decrease in response rates, break-offs, and contain more missing values (e.g., Dillman, Sinclair, and Clark, 1993). Furthermore, long questionnaires tend to lose respondents who are not very involved or interested in the investigated domain, which eventually may lead to a sampling bias (Moore et al., 2002). From a practical perspective, long measurement instruments are more costly because additional expenses for data collection and coding need to be covered (Moore et al., 2002).[1] Overall, since space on a questionnaire is limited and therefore very valuable, researchers frequently face pressure when deciding which constructs to include or to exclude in a study; all other things being equal, the more constructs that are included, the greater the demands on measurement in terms of number of items to be included in the survey questionnaire.

Rather than omit potentially relevant constructs (and hence risk misspecification of theoretical models), researchers may opt for the use of shortened versions of existing multi-item scales or even employ single-item measures to operationalize (some of) their constructs. Indeed, several researchers in psychology have highlighted the need of making measurement more efficient by either substantially reducing the length of the scales applied or by using single-item measures as opposed to multiple item measures (e.g., Stanton et al., 2001; Nagy, 2002; Russell et al, 2004). With specific reference to the latter, single-item measures offer the important advantages of being short, flexible and easy to administer (Pomeroy, Clark and Philip, 2001). They are also less time consuming and not monotonous to complete (Gardner et al., 1998), thus reducing response biases (Drolet and Morrison, 2001). However, conventional measurement wisdom in business and management research strongly advocates the use of multi-item scales (e.g., see Boyd, Gove and Hyatt, 2005) and, indeed, practically *all* measure development textbooks (e.g., DeVellis, 2003; Netemeyer, Bearden and Sharma, 2003; Viswanathan, 2005) and articles (e.g., Churchill, 1979; Comrey, 1988; Homburg and Giering, 1996; Diamantopoulos and Winklhofer, 2001) focus on the construction of multi-item measures. Accordingly, editors and reviewers of academic journals are reluctant to accept manuscripts using single-item measures to operationalize some of their constructs (Singh, 2003); in fact, »the use of single-item measures in academic research is often considered a »fatal error« in the review process« (Wanous, Reichers and Hudy, 1997, p. 247).

Recently, however, several authors in the management field have challenged such conventional wisdom by demonstrating that single-item measures *can* have acceptable psychometric properties and are, therefore, a potentially viable alternative to multi-item scales for construct measurement purposes (e.g., Drolet and Morrison, 2001; Bergkvist and Rossiter, 2007). Consequently, »the use of single-item measures should not be considered fatal flaws in the review process. Rather, their appropriateness for a *particular* piece of research should be *evaluated*« (Wanous, Reichers and Hudy, 1997, pp. 250–251, added emphasis). Unfortunately, there is currently little guidance to practising researchers in the management field regarding the circumstances under which the use of single-item measures may be acceptable. This is both because most methodological work on single-item measures has been conducted outside the management discipline and because different authors have focused on specific aspects of single-item measures (e.g., reliability or predictive validity) in isolation. What is missing is an *integrative* framework within which the potential acceptability of single-item measures can be comprehensively evaluated on multiple relevant criteria.

It is this gap in the literature that the present paper seeks to fill. Specifically, drawing from an interdisciplinary literature review on the use of single-item measures, we aim to identify the con-

1 In most instances the cost of the survey depends on the number of questions asked (Brace, 2004).

ditions under which their use can be justified in substantive empirical research. Our objective is to provide management researchers with concrete guidelines on how to assess the extent to which a single-item measure can be legitimately used to operationalize the focal construct.

In the section that follows, we provide a conceptual perspective on single-item measures and follow this with a discussion of reliability and validity assessment issues. Next, we present key considerations influencing the choice of single- vs. multi-item measures and conclude the paper with some suggestions regarding future research on single-item measures. Throughout the following discussion, a single-item measure is considered to be an individual measure or indicator as defined under the »total disaggregation« model of construct representation by Bagozzi and Heatherton (1994). In other words, a single-item measure, as conceived here, *cannot* be further decomposed to lower-level constituents as is the case, for example, with composite indicators made up of several items which are aggregated/averaged to compute the composite. An illustrative example of a single-item measures is »All in all, would you say that you are satisfied or dissatisfied with your job?«, scored on, say, a 7-point scale ranging from »very dissatisfied« to »very satisfied«.

2. A Conceptual Perspective on Single-Item Measures

2.1. Reflective Measurement

Given a focal construct η and a measure of it x_1, a single-item operationalization under a reflective measurement perspective can be described as follows:

$$x_1 = \lambda_1\eta + \varepsilon_1 \quad (1)$$

where λ_1 is the loading of x_1 on η and ε_1 is the measurement error associated with x_1. It is assumed that $COV(\eta, \varepsilon_1) = 0$ and that $E(\varepsilon_1) = 0$. Note that equation (1) assumes that x_1 is a unidimensional item, that is, only a *single* latent variable (i.e., η) contributes to its variation.[2]

There are two ways one can conceptually approach equation (1). The first way assumes that x_1 is the *only* measure that can be used to represent η or, what amounts to the same thing, that changes in η are solely and exclusively reflected in changes in x_1. Here one faces the problem that it is not possible to estimate the parameters λ_1 and θ_1 (the variance of ε_1) because the model in equation (1) is underidentified.[3] The usual »fix« in this case is to either assume that x_1 is a perfect measure of η (which implies an error variance of zero, i.e., $\theta_1 = 0$) or to assume a certain level of reliability for x_1 and use this to set the error variance to a specific value (Jöreskog and Sörbom, 1989); for example, if a reliability of .80 is assumed, θ_1 would be constrained to $(1-.80) \times VAR(x_1)$.

Neither of these alternatives is particularly attractive from a conceptual perspective. Assuming complete absence of measurement error implies that »a concept becomes its measure and has no theoretical meaning beyond that measure« (Bagozzi, 1982, p. 15). Assuming a certain amount of measurement error is a better option but still unsatisfactory because there is no way of knowing/testing whether the chosen value is accurate or not. Given that a measure's reliability (or lack of) impacts its relationships with other measures (see Nunnally and Bernstein, 1994), depending on the reliability values chosen to fix θ_1, conclusions about the substantive linkages between η and other constructs may well differ.

A second problem arises because it is difficult to provide a convincing answer to the question *why* is x_1 the *only* indicator that can be used to operationalize η, or put differently, why *no other* indicator(s) can be possibly used. In this context, »an observable measure never fully exhausts everything that is meant by a construct … a measure of a construct could never have complete content validity« (Peter, 1981, p. 134). Why cannot a second indicator, say x_2, also capture the part of the meaning of η that is already captured by x_1 and/or that part which is *not* captured by x_1? In other words, what is it that makes x_1 *unique*? A possible answer to this question is that it is sometimes »difficult to generate

2 We focus initially on the unidimensional case, because of the acknowledged desirability of unidimensionality as a measurement property in management research (e.g., see Anderson and Gerbing, 1982; Gerbing and Anderson, 1988; Danes and Mann, 1984; Steenkamp and van Trijp, 1991). We will consider the case of multiple latent variables impacting on x_1 later in this section.

3 Assuming that the variance of η is standardized, there are two parameters to be estimated, namely λ_1 and $\theta_1 = VAR(\varepsilon_1)$, but only one piece of information (the variance of x_1).

multiple items to measure a construct. The major reason is that the construct is simple and single-faceted, and it is *impossible* to create many different items that measure the same underlying construct« (Poon, Leung and Lee, 2002, p. 276, added emphasis). Indeed, within the context of the C-OAR-SE scale development procedure, Rossiter (2002, p.331) argues that »a concrete singular object to be rated in terms of a concrete attribute needs only a single-item scale«.

An alternative way one can approach equation (1) is by considering x_1 as one of a set of indicators that could be *potentially* used to measure η. Here, x_1 is not seen as *the* measure of η but as a *representative* measure of η. This viewpoint is much more consistent with the domain sampling model in measurement theory which considers »any particular measure to be composed of responses to a random sample of items from a hypothetical domain of items« (Nunnally and Bernstein, 1994, p. 216). A single-item measure such as x_1 can thus be considered to be *one* possible measure of η drawn from the relevant domain. Bearing in mind that »one major source of measurement error occurs when the sampling of the domain items is inadequate« (Grapentine, 2001, p.156), one can ask the question whether a single item can ensure adequate domain representation. This is essentially a sampling question and according to sampling theory (e.g., see Sudman, 1976; Cochran, 1977), a key consideration is the extent of variability/heterogeneity in the target population; all other things being equal, the more homogeneous the population, the smaller the needed sample size.[4] In the extreme, if all population elements are identical to each other, then a sample of one is adequate to accurately represent the population.

Applying this logic to the domain sampling model, x_1 could be considered as being representative of the domain of relevant items if the other items were similarly related to η as x_1. This applies in the case of parallel items, whereby »each item of a scale is precisely as good a measure of the latent variable as any other of the scale items... each item's relationship to the latent variable is presumed to be identical to the every other item's relationship to that variable *and* the amount of error present in each item is also presumed to be identical« (DeVellis, 2003, p.21, original emphasis). If one were to assume that x_1 is one of several parallel items, it becomes possible to estimate the parameters in equation (1) as long as a second item, x_2, is available. The measurement equation for that second item would be

$$x_2 = \lambda_2 \eta + \varepsilon_2 \qquad (2)$$

with λ_2 representing the loading of x_2 on η and ε_2 measurement error.[5] Given that under the assumption of parallelism $\lambda^2_1 = \lambda^2_2$ and $\theta_1 = \theta_2$, only *two* parameters need to be estimated (i.e., one loading and one error variance) from *three* pieces of information (i.e., the variance of x_1, the variance of x_2, and the covariance between x_1 and x_2). Thus the measurement model defined by equations (1) and (2) becomes overidentified (with 1 degree of freedom) and can be estimated/tested. Subsequently, either x_1 or x_2 could be used as a single-item measure of η in a larger model (including antecedents and/or consequences of η) with the error variance of the chosen indicator (i.e., either θ_1 or θ_2) fixed at the value obtained during the estimation of the parallel model. Note, in this context, that x_1 and x_2 (and, for that matter, all other parallel items to x_1 and x_2) are perfectly interchangeable and empirically indistinguishable and can only differ in terms of their specific wording (i.e., their measurement properties are identical). Of course, one runs the (very real) danger that x_1 and x_2 are merely semantically identical (i.e., redundant) items, whereby »essentially the same item is rephrased in several different ways« (Boyle, 1991, p. 281). In this case, »the second item in such a pair does not represent additional sampling from the content domain, and so one aspect of the domain may be oversampled« (Smith and McCarthy, 1995, p. 306).[6]

One must also pose the question whether the assumption of parallel items is too restrictive when operationalizing constructs in management research. The measurement tradition in this research field has been Jöreskog's (1971) congeneric model which merely assumes »that all the items share a common latent variable. They need not bear equally strong relationships to the latent variable, and their error variances need not be equal« (DeVellis, 2003, p. 25). Thus, unlike with parallel items, under con-

4 Given fixed confidence and precision levels.

5 The usual assumptions $COV(\eta, \varepsilon_2) = 0$ and $E(\varepsilon_2) = 0$ also apply.

6 Thus it is recommended to »retain items that are parallel but not those that are identical« (Smith and McCarthy (1995), p. 306).

generic measurement assumptions, items are *not* identical to each other as far as their measurement properties are concerned. This complicates the definition of a »representative« item to be used as a single-item measure of the focal construct. This is a *conceptual* problem which is not easily resolved even if one has previously developed a psychometrically-sound, multi-item scale following established measure development procedures (e.g., Churchill, 1979; Spector, 1992; Netemeyer, Bearden and Sharma, 2003). For example, choosing the item with the highest reliability[7] or the item with highest loading (as done, for example, by Loo, 2002) does not conceptually explain *why* such an item is more representative of the construct domain than the other items (thus making it more suitable as a candidate for a single-item measure).[8] Moreover, what may prove to be the »best« item judged on the above criteria in one sample may well turn out to be not the »best« one in a different sample. Purely relying on empirical results to identify the »best« item is bound to capitalize on the idiosyncrasies of the sample data at hand and, short of replicating the full scale on several samples, there does not appear to be an easy way out of this problem.

Finally, if the unidimensionality assumption is relaxed, then a single-item reflective measurement model becomes particularly problematic. Consider model shown in equation (3). Here, variation in x_1 is attributable to *two* latent variables η_1 and η_2, which means that x_1 measures two constructs rather than a single construct as was assumed so far.[9] In the absence of any other measures for η_1 and η_2, it is not possible to identify *which* of the latent variables is responsible for the covariation with the other variables in the system. In addition to the obvious identification problems of the model shown in equation (1), there is a major conceptual difficulty with accepting that a *single* item can be a *sufficient* measure for *multiple* constructs.[10]

$$x_1 = \lambda_1 \eta_1 + \lambda_2 \eta_2 + \varepsilon_1 \qquad (3)$$

2.2. Formative Measurement

A single-item operationalization under a formative perspective can be described as follows:

$$\eta = \gamma_1 x_1 + \zeta \qquad (4)$$

where γ_1 is the expected effect of x_1 on η and ζ is a disturbance term. It is assumed that $COV(x_1, \zeta) = 0$ and $E(\zeta) = 0$.

Unlike with the reflective measurement perspective discussed in the previous section, there is only one interpretation of equation (4), namely that η is determined by x_1 as well as by ζ_1, the latter capturing »the impact of all remaining causes other than those represented by the indicators included in the model« (Diamantopoulos, 2006, p. 11). Note that x_1 *cannot* be conceived as a »representative« indicator from a larger set of possible items because formative measurement »does not follow the domain *sampling* model. This means that the items are *not* interchangeable« (Rossiter, 2002, p. 315, original emphasis).[11]

Bearing the above in mind, a single-item measure under a formative perspective appears problematic for several reasons. First, because »the latent variable is determined by its indicators rather than vice-versa, content specification of the construct] is in extricable linked with indicator specification« (Diamantopoulos and Winklhofer, 2001, p. 271). This implies that a *census* of indicators is needed (Bollen and Lennox, 1991). It is difficult to convincingly argue that such a census has been achieved by just using a single item. Second, as MacKenzie, Podsakoff and Jarvis (2005, p. 727) point out, »it is not conceptually possible for a formative indicator to perfectly represent a composite latent construct because the construct is defined as function of multiple distinct components

7 As indicated, for example, by the item's squared multiple correlation (SMC) in a LISREL analysis (see Bollen, 1989).

8 In fact if the chosen item has *much* higher reliability and/or loading than the rest of the items, one could argue that it would *not* be representative of the other items!

9 It is assumed that $E(\varepsilon_1) = 0$ and $COV(\eta_1, \varepsilon_1) = COV(\eta_2, \varepsilon_2) = 0$. For measurement models with multidimensional items, see Bollen (1989, pp. 198–199).

10 The authors would like to thank an anonymous reviewer for suggesting expanding the discussion of single-item reflective measures to models where item unidimensionality is not assumed.

11 Note that, like all formative models, the model shown in equation (4) is, on its own, statistically underidentified. Note also that addition of more items does *not* overcome the identification problem and »such models are unidentified unless embedded within a larger model which includes additional constructs measured with reflective indicators... or unless additional reflective items are included as consequences of the formative construct« (Diamantopoulos, 2006). An excellent discussion of identification issues in the context of formative models can be found in Temme (2006).

or parts, so one indicator cannot validly represent the entire conceptual domain«. Third, with a single indicator, it is more than likely that the variance of the residual (VAR(ζ) in equation (4)) will be substantial (Diamantopoulos, 2006). In this context, »as the residual increases, the meaning of the construct becomes progressively ambiguous... Certainly, the meaning of a construct is elusive when most of its variance is attributable to unknown factors« (Williams, Edwards and Vandenberg, 2003, p. 908).[12]

There is, however, one way one can conceive a single-item measure within a formative measurement perspective, namely, to specify »a global item that summarizes the essence of the construct that the formative] index purports to measure« (Diamantopoulos and Winklhofer, 2001, p. 272). Note that this item is *not* part of the set of formative indicators used to specify the focal construct but is typically used as an initial check of the validity of the individual indicators (e.g., see Diamantopoulos and Siguaw, 2006). A good example of this approach can be found in the job satisfaction literature where »global« or »overall« single-item measures of satisfaction are compared with multi-item scales capturing different *facets* of job satisfaction (e.g., see Wanous, Reichers and Hudy, 1997; Oshagbemi, 1999; and Dolbier et al, 2005 and references therein). A theoretical rationale for using such global job satisfaction measures is that »a worker would be generally satisfied if s/he is satisfied in all important facets of job satisfaction (e.g., satisfaction with pay, with supervisor, with co-workers, with promotion and with work itself...). In other words, the facet satisfaction leads to overall satisfaction« (Law and Wong, 1999, p. 149). Such a »global« item approach to specifying single-item measures has been used in several other research fields, including quality of life assessment (e.g., de Boer et al, 2004), sport management (e.g., Kwon and Trail, 2005), citizen satisfaction (e.g., Van Ryzin, 2004), self esteem (Robins, Hendin and Trzesniewski, 2001), and teaching effectiveness assessment (e.g., Wanous and Hudy, 2001).

It is important to note, however, that none of the aforementioned studies (either in the job satisfaction or other research fields noted above) *explicitly* used a formative perspective as a theoretical foundation when specifying »global« or »overall« single-item measures. The latter have been invariably developed on an *ad hoc* basis and, in fact, the multi-item measures with which they have been compared are typically *reflective* scales! This raises issues of measurement model misspecification for several of the constructs examined (job satisfaction being a prime example).[13] Although such misspecification issues are clearly beyond the scope of the present paper, it is important to bear in mind that they *do* exist and suggest caution in interpreting the relevant literature.[14]

In conclusion, there are several thorny issues – primarily of a conceptual nature – associated with the specification and selection of single-item measures under both a reflective and formative measurement perspective. Unfortunately, these issues are hardly ever addressed by proponents of single-item measures; Rossiter (2002, p. 314, added emphasis), for example, states that »the goal is to develop *one good* item for each first-order component »(meaning each facet of a formatively-measured construct). However, he offers no insights as to exactly *what* makes a good item (which is clearly what the real issue is all about). In this context, the typical approach in relevant methodological research has been to *infer* the quality of a single-item measure (usually specified on a more or less ad hoc basis) by comparing its relative performance with an established (and psychometrically sound) multi-item scale of the focal construct in terms of reliability and/or predictive validity. The underlying rationale is well-described by Gardner et al (1998, p. 899) who state that »it is possible that one »good« item can be better than many »bad« items when evaluated on criteria of reliability and validity«. It is to a discussion of such issues that we now turn.

3. Reliability Assessment

Reliability, in the context of measurement, is defined as »the degree to which measures are free from error and therefore yield consistent results«

12 Note also that »it is not empirically possible for one indicator to account for all of the variance in the composite latent construct unless the indicators are perfectly correlated (which is unlikely to be the case)« (Mackenzie, Podsakoff and Jarvis, 2005, p. 727).

13 For a detailed exposition of why current conceptualizations of job satisfaction should be operationalized under a formative perspective, see Law and Wong (1999).

14 For a review of the potential consequences of measurement model misspecification, see Diamantopoulos, Riefler and Roth (2008) and references given therein.

(Peter, 1979, p. 6). Reliability constitutes a crucial psychometric premise of any measure (Spector, 1992; Netemeyer, Bearden and Sharma, 2003). It is commonly agreed that longer scales are more reliable than short scales (Nunnally and Bernstein, 1994; Spector, 1992; Lowenthal, 2001; Rust and Golombok, 1989).[15] Single-item measures in particular, are deemed to be »notoriously unreliable« (Spector, 1992, p. 4) both because measurement errors of the individual items are not smoothed out by the summation of the item-scores to a total score and because inconsistent responses may be obtained in successive administrations of the research instrument (Churchill, 1979; Epstein 1979). On the other hand, there is evidence to suggest that »additional items can significantly inflate error term correlation and thereby reduce the informal value of added items« (Drolet and Morrison, 2001, p. 200) meaning that the *incremental* information provided by each additional scale item is extremely small. Specifically, if items are semantically similar, respondents tend to make inferences from the content of one item to the remaining items of the scale, assuming that they are basically the same without carefully reading them. In general, respondents who are exposed to more items tend to distinguish less between them, with earlier items having a stronger influence on later items; hence more items may lead to mindless response behaviour (Drolet and Morrison, 2001). A second related problem is that multiple items are prone to consistency motif bias, »in which subjects tend to try to maintain consistency in their responses of similar questions« (Podsakoff et al., 2003, p. 881).

However, whether a single-item or multi-item measure is more reliable is, ultimately, an empirical question. Concerning test-retest reliability, there is evidence suggesting that single-items can be very reliable (e.g., see de Boer et al, 2004; Shamir and Kark, 2004). As far as internal consistency reliability is concerned, again, there is substantial evidence indicating acceptable reliability values for single-item scales (e.g., see Ginns and Barrie, 2004; Wanous, Reichers and Hudy, 1997; Wanous and Hudy, 2001; Kwon and Trail, 2005; Jordan and Turner, 2005; Dolbier et al, 2005).[16] Note, in this context, that »it is frequently said that one cannot estimate the internal consistency reliability of single-item measures and this alone is sometimes believed to be a sufficient reason to limit or avoid their use« (Wanous, Reichers and Hudy, 1997, p. 247). This misconception is particularly prevalent among management researchers as illustrated by the following passage from a review paper in the highly prestigious *Strategic Management Journal*:

»*Single measures* at the nadir of methodological sophistication, provide the researcher with the least assurance that a measure is a valid and reliable proxy of a construct and no estimates of reliability, and thus error, are possible« (Boyd, Gove and Hitt, 2005, p. 244)«.

Such misconceptions can probably be traced to the fact that, in structural equation models with reflective indicators, no parameter estimation can take place when only a single indicator is used to operationalize the construct of interest; this, however, is simply because the relevant measurement model is underidentified (see earlier discussion of equation (1)) and it makes no difference as to *which* particular indicator is used in the model. When a single-item measure is contrasted with a multi-item measure, however, there are several ways of establishing the reliability of the former based on such techniques as factor analysis and the correction for attenuation formula (see Wanous and Reichers, 1996 and Wanous and Hudy, 2001 for formulae and illustrative applications). Moreover, assuming that one wants to select a single item from a scale of k items, one can apply the Spearman-Brown prophecy formula in reverse (see Nunnally and Bernstein, 1994, p. 263–264), whereby the reliability of a scale »$1/k$ th« as long as the original scale would be estimated.

In summary, an outright rejection of single-item measures on reliability considerations does not seem to be justified. One *can* estimate the reliability of single-item measures and the resultant estimates are, more often than not, within acceptable levels. However, »reliability is a necessary but not sufficient condition for validity« (Churchill, 1979, p. 65; see also Peter, 1979). A discussion of validity

15 Note, however that »an increase in the number of the scale items leads to participant fatigue, boredom, and inattention, which, in turn, can lead to inappropriate (mindless) response behaviour« (Drolet and Morrison, 2001, p. 198). Hence, there is a trade off between the parsimony of a scale and its reliability (DeVellis, 2003).

16 Of course, there is also evidence to the contrary, that is, indicating insufficient reliability of single-item scales (e.g., see Loo, 2002; Epstein, 1979).

considerations as applied to single-item measures follows.

4. Validity Assessment

Construct validity refers to »the degree to which a measure assesses the construct it is purported to assess ... In this sense a measure is construct valid (1) to the degree that it assesses the magnitude and direction of representative sample of the characteristics of the construct and (2) to the degree that the measure is not contaminated with elements from the domain of other constructs and error« (Peter, 1981, p. 134).

It is asserted that single-item measures lack validity, because they tend to insufficiently capture the conceptual domain of most constructs (Nunnally and Bernstein, 1994); thus compared to multi-item measures, they do not tap into a construct from different angles (Baumgartner and Homburg, 1996; Wirtz and Lee, 2003). On the other hand, these arguments are traded off by the (common) »practice of adding attempted synonyms that actually decrease the content validity of the measure« (Rossiter, 2002, p 331). This problem is confirmed by Hinkin (1995), who, in a review of scale development practices in the field of management, ascertained that long scales possessed sound reliabilities but frequently picked up substance from more than one conceptual domain. If this is indeed the case, then obviously »one or two *good* items that elicit *appropriate* respondent behaviour will yield better *information* than multiple, poorly presented items that increase the error term correlations and/or stimulate inappropriate response styles« (Drolet and Morrison, 2001, p. 199, original emphasis). A related point, from a face validity perspective, is that single-item measures may be advantageous because »respondents may resent being asked questions that appear to be repetitions... From a management perspective, a single-item is usually easier to understand than a scale score« (Wanous, Reichers and Hudy, 1997, p. 250).

A particular concern with single-item scales concerns the assessment of convergent and discriminant validity because »each item tends to relate to attributes other than the one to be measured« (Nunnally and Bernstein, 1994, p. 66). Given two constructs η_1 and η_2 each measured by a single indicator, say, x_1 and x_2 respectively and assuming standardization of the latent variables, five parameters need to be estimated, namely λ_{11}, λ_{22}, $\theta_1 = VAR(\varepsilon_1)$, $\theta_2 = VAR(\varepsilon_2)$ and $\Phi_{12} = COV(\eta_1, \eta_2)$.[17] As there are only three pieces of information available (i.e., $VAR(x_1)$, $VAR(x_2)$ and $COV(x_1, x_2)$), the model is underidentified and, therefore, no estimates of convergent validity (as captured by the indicator loadings λ_1 and λ_2) or of discriminant validity (as captured by the magnitude of the interconstruct correlation, Φ) can be obtained. Thus when both constructs are measured with just one indicator each, it is not possible to either distinguish between the constructs or test whether each indicator indeed loads stronger on its posited construct rather than on a different construct. For instance, the constructs of »brand attitude« and »brand satisfaction« would not be formally distinguishable when single-item measures are used.

In light of the above, the common approach for assessing convergent validity has been the computation of the correlation between the single-item measure and its multi-item counterpart (i.e., the »full« measure of the construct). The empirical evidence based on this approach has been, overall, encouraging for single-item measures. For example, in a job satisfaction context, Wanous, Reichers and Hudy's (1997) meta-analytic study reported a mean correlation between single items and full scales of .63 and a corrected correlation of .67.[18] Another meta-analytic study of teaching effectiveness revealed mean observed and corrected correlations of .79 and .84 respectively (Wanous and Hudy, 2001). Similarly encouraging results have been reported by Robins, Heudin and Trzesniewski (2001), Nagy (2002), and Dolbier et al. (2005), among others. Although the discriminant validity of single-item measures has been investigated less often, those (few) studies who have done so also reported findings in support of single-item measures (e.g., see Gardner et al, 1998, and de Boer et al., 2004).

Finally, several studies have compared the predictive validity of single-item measures to that of multi-item scales. With few exceptions, the evidence from studies in fields as diverse as health care (DeSalvo et al, 2006), sports management (Kwon and Trail, 2005), organizational psychology

17 Given that η_1 and η_2 are standardized, Φ is a correlation coefficient.

18 Corrected for unreliability.

(Nagy, 2002) and marketing (Bergkvist and Rossiter, 2007) shows that single-item scales can have good predictive validity (comparable to those of their multi-item equivalents).

Taken collectively, the findings relating to the validity of single-item measures fail »to support the classic psychometric argument (e.g., Churchill, 1979; Nunnally and Bernstein, 1994) that multiple-item measures are more valid than single-item measures for all types of constructs« (Bergkvist and Rossiter, 2007, p.182). Thus single-item measures can be both reliable and valid, implying that their potential application should not be ruled out on a priori grounds.

5. Criteria for the Use of Single-Item Measures

In the previous two sections it was shown that, contrary to common beliefs, single-item measures can have acceptable psychometric properties. In this context, it has been argued that »if single-item measures can be shown to adequately and accurately represent the variables being measured, then many researchers ... could benefit. There are several advantages to using single-item scales: simplicity, brevity or ease of use, and global measurement« (Kwon and Trail, 2005, p. 72). We therefore now establish potential criteria that should assist researchers decide whether the use of single-item measures is acceptable or not. In doing so, we identify conditions or situations, in which the use of single items would be worth serious consideration. The focal criteria can be summarized into four broad groups corresponding to (1) the nature of the construct, (2) the nature of existing instruments, (3) the research objectives, and (4) sampling considerations.

5.1. Nature of the Construct

The selection of single- versus multiple-items depends to a great extent on the construct of interest. Particularly relevant in this respect is whether the focal construct is concrete or abstract (Rossiter, 2002). Concrete constructs refer to objects and their characteristics which are perceived similarly by all raters, whereby there is »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). Examples are favorability, price perception, and buying intention.

Abstract constructs, on the other hand, mean different things to different raters and thus, are perceived as heterogeneous by raters; examples are creativity, power, or corporate culture. It is generally held that, with abstract constructs, the use of multiple-item measures is required, because »most constructs, by definition, are too complex to be measured effectively with a single item ...« (Peter, 1979, p. 16). On the other hand, when a construct is judged to be concrete, the use of single item measures is considered reasonable (Sackett and Larson, 1990; Rossiter, 2002), not least because measurement error is more prevalent for abstract versus concrete concepts (Cote and Buckley, 1987). We basically share this view and advocate the use of single-item measures for investigating concrete constructs.

Related to the concreteness of the construct is its complexity/dimensionality. Generally, for multidimensional constructs, the use of single-item measures is inappropriate (Drolet and Morrison, 2001b, Nunnally and Bernstein, 1994). According to Loo (2002, p. 73), for example, »single-item measures may be considered only if the single item reflects a homogeneous construct, as indicated by a high internal consistency reliability coefficient (alpha > 0.85) or a unidimensional construct as reflected by an item factor analysis.« Hence, the use of single-item measures is appropriate if the construct of interest is unidimensional rather than multidimensional (Sackett and Larson, 1990; Nunnally, 1978).[19] However, in cases in which a construct is multidimensional and its respective dimensions are known and have been empirically established, *each* dimension may be measured with a single item, *provided* that the latter is reliable.

While it is usually the case that with increasing complexity of the construct (i.e., increasing number of dimensions), additional questions (and hence items) need to be asked to cover all dimensions of the latter, »there comes a point where the construct becomes so complex that a single question may be

19 See also the discussion in Section 2.2 regarding item dimensionality.

the best.« (Sloan et al., 2002, p. 481). With *highly* complex constructs, it may be that not all dimensions of the latter are covered by traditional multi-item measures, thus, resulting in incomplete evaluations of the construct (Nagy, 2002). Put differently, in order to cover each and every potential aspect of a highly complex construct, an enormous number of items would be needed that would render a practical application impossible. As a result, Scarpello and Campbell (1983), for example, contend that the best way to measure overall job satisfaction is to use a single question along the lines »Overall, how satisfied are you with your job?«

When asking »global« single-item questions such as the above, it is assumed that respondents »automatically« consider different aspects of the construct (provided that respondents are aware of the full scope of the concept). Single-item measures allow a respondent to »consider all aspects and individual preferences of the certain aspects of the construct being measured« (Nagy, 2002, p. 79), and thus provide a more »tailor-made« picture of that respondent's construct view.[20] Consequently, with »global« single-item measures, respondents tend to ignore aspects that are not relevant to their situations, and differentially weight the relevant aspects to provide a single rating (De Boer et al., 2004); in contrast, multiple-item scales employ external subjective or statistical weighting to combine items to come to an overall rating. In this context, the question arises whether it is better if researchers decide which facets of the concept are to be measured, and weigh them based on statistical considerations or if consumers should be empowered to decide which facets are important or not important. For example, if consumers are asked for their quality of life status, they may consider more facets or alternative facets which may be of particular importance to them, compared to a fixed number of facets included in standardized multiple-item measures.

Needless to say that constructs must be accurately described and made clear to respondents when measured with a single-item measure (Sackett and Larson, 1990). This is because single-item measures require more abstract thinking as opposed to multiple-item scales (Sloan et al., 2002) and, therefore, may be too vague for respondents to be »correctly« answered.[21] Especially, the use of »global« single-item measures, which usually capture constructs that are *not* concrete, requires that respondents are aware of every potential facet of the construct; otherwise, ambiguous interpretations may result which, in turn, will adversely impact the quality of the responses obtained. Thus, regarding highly complex constructs, we recommend that single-item measures are used with great care, paying particular attention to the degree of the respondents' understanding of the focal construct.

5.2. Nature of Existing Instruments

As already mentioned in previous sections, in an effort to maximize the internal consistency of scales, numerous constructs in business and management research possess semantically identical and therefore redundant items (Albers and Hildebrandt, 2006), whereby »essentially the same item is rephrased in several different ways« (Boyle, 1991, p. 281). With semantically identical items one aspect of the domain is likely to be oversampled (Smith and McCarthy, 1995). This tends to lead to the effect that the same error variance associated with the particular item is also likely to be associated with the other items (Drolet and Morrison, 2001, Smith and McCarthy, 1995); as a result, measurement error does *not* cancel out in the aggregation process, which negatively affects reliability and precision.

Redundant items also contribute to the known negative effects of lengthy scales, namely boredom, fatigue, and various response biases (Duhachek, Coughlan, and Iacobucci, 2005; Drolet and Morrison, 2001). Hence, we advise researchers to carefully scrutinize existing scales and determine the extent to which their items are semantically identical/similar; this judgment should be conducted by at least two experts (coders) independently to ensure a high degree of objectivity. Very high estimates of item homogeneity (as reflected in inter-item correlations) are indicative of potential

20 Thus single-item measures »allow the subject to take personally salient features of the situation into account when providing a response« (Youngblut and Casper, 1993, p. 459).

21 As pointed out by Sloan et al., 2002, p. 484, »the level of abstract thinking required of the respondent may be greater when using single-item measures than when using multi-item indices: For example, it may be more cognitively challenging to provide an overall rating of one's level of social functioning than to respond to a series of relatively concrete questions about spending time with family and friends«.

item redundancy (Boyle, 1991). Note, however, that high inter-item correlations should not be *automatically* equated with semantically redundant items, because items measuring different aspects of a construct may be highly related. However, in case where a high degree of redundancy among items *is* established, we subscribe to Drolet and Morrison's (2001, p.197) view that »information content can be negatively affected by scales with multiple items« and, therefore, strongly recommend the use of a single-item measure.

5.3. Research Objectives

The issue whether to use a single-item measure also depends on the respective objective(s) of the study, and, in particular, the role of the construct in the research design (Sloan et al., 2002). If a construct is at the heart of a study, with the researcher's intention to generate specific insights into the nature of the construct, a more detailed measurement approach applying full-length multiple-item scales should be chosen. The latter are superior over single-item measures for identifying detailed aspects or highlighting certain characteristics of constructs (Lee et al., 2000). If, on the other hand, investigators are only interested in obtaining a general view of the construct and the research objective is to get an overall feeling, judgment, or impression on the latter, a single-item measure is often adequate for the purpose (Poon, Leung and Lee, 2002). Indeed, »the single-item global rating method may be useful if the goal of a study is to gain an understanding for the general nature of aconstruct]« (Lee et al., 2000, p. 242). Finally, if the construct is only of secondary importance in the study setting (e.g., is used as moderator, validation, or control variable), the employment of single-item measures can be justifiable.[22]

In contrast to multi-item measures, single-item measures are criticized for lacking precision because they have a tendency to categorize people into a relatively small number of groups (e.g. Spector, 1992; DeVellis, 2003; Nunnally and Bernstein, 1994; Netemeyer, Bearden and Sharma, 2003; Churchill, 1979). Hence, multiple item measures possess more responsiveness, that is, the ability of a measure to detect small but important differences (de Boer et al., 2004).[23] This is crucial, because many measurement problems require a very fine differentiation or categorization (Nunnally and Bernstein, 1994) and some measures are purposely constructed with the aim to clearly differentiate among respondents or to differentiate extreme cases from less extreme ones (Smith and McCarthy, 1995). Having said all that, the inclusion of more points in a single-item measure may be a (partial) remedy for the precision issue (Bergkvist and Rossiter, 2007). For example, according to Wirtz and Lee (2003), more finely grained single-item measures are able to discriminate between different degrees of the measure similarly to multiple-item measures.

In longitudinal studies, the use of single-item measures is often recommended over multi-item measures because regularly completing a lengthy scale (or questionnaire) puts a big burden on the respondent (Drolet and Morrison, 2001; Nagy, 2002; Wanous, Reichers and Hudy, 1997). However, although well-meant, such recommendations may be misguided because with a single-item measure, it is not possible to differentiate between (a) »true« changes in the underlying construct over time, (b) reliability of measurement, and (c) situational factors that may impact responses when measurement takes place at different points in time (as is the case in longitudinal studies). Indeed, according to latent state-trait theory,[24] »measurement does not take place in a situational vacuum; a person can be measured only in a situation which might have a systematic effect on each variable measured at occasion k ... Several instruments measuring the same state on each occasion are necessary in order to obtain information about the reliability of the measurements. Repeated observations on *several* occasions are necessary in order to get information on the degree of consistency over interindividual differences over time« (Steyer and Schmitt, 1990, p. 433 and p. 437, emphasis in original).

In light of the above, we advise against the use of single-item measures in longitudinal research

22 For example, in a strategic management setting, Boyd, Gove and Hitt (2005) show that control variables are nearly exclusively captured using single items (79.8% of control variables); besides, 38.1% of dependent variables and 47.6% of independent variables used single-item indicators.

23 In this context, the relatively large number of total scores makes it possible to »make relatively fine distinctions among people (Churchill, 1979, p. 66).

24 For a discussion of latent state-trait models and applications, see Steyer and Schmitt (1990), Steyer, Ferring and Schmitt (1992) and Steyer, Schmitt and Eid (1999).

and subscribe to Steyer's Schmitt and Eid's (1999, p. 403) recommendation that »we need repeated measurements on at least two occasions with at least two instruments measuring the same states«.

5.4. Sampling considerations

If a measure is to be administered to a wide range of different populations, the use of single-item measures has certain advantages. Indeed, an »advantage of the single-item or short scales is that they can be given to numerous people« (Gorsuch and McPherson, 1989, p. 352). Indeed, one major strength of single-item measures is their flexibility; for example, »whether one is investigating the job satisfaction of pilots, fishermen, solicitors or managers, the single-item measure of job satisfaction may be administered.« (Oshagbemi, 1999, p. 393).

In the field of organizational research, managers frequently feel that they are »over-surveyed«, which contributes to the issue of low response rates (Rogelberg and Luong, 1998). Specifically, the difficulty of attaining large sample sizes in surveys, due to lack of willingness to sacrifice time to complete questionnaires, leads to the necessity of reducing the length of construct measures. As a rule of thumb, there should be at least ten times as many respondents as items or, in cases where a large number of items are used, at least five respondents per item (Nunnally, 1967; Peter, 1979). Taking these considerations into account, we suggest that if only a small sample size is available (e.g., due to budget constraints, difficulties in recruiting respondents, requirement of dyadic data, and other similar constraints), considering the use of single-item measures may be a pragmatic solution.

Table 1 summarizes the criteria along which researchers may judge whether measurement with a single item is acceptable/justifiable. The more criteria that are fulfilled, the more acceptable the use of single-item measures. For instance, if a construct is concrete, used as a control variable, and the single-item measure sufficiently satisfies reliability and validity requirements, it would indeed be appropriate to use it in substantive research.

6. Conclusions

This article contributes to the management literature by providing insights into the conceptual nature of single-item measures and by dispelling several myths regarding their psychometric properties. Specifically, our analysis shows that the application of single-item measures is appropriate under certain conditions and that their general banishment is not justified. By considering the criteria offered in this paper, researchers should be able to reach an informed decision as to whether the use of a single-item measure for the focal construct is acceptable given the research purpose and setting at hand. Assuming that it is, it is important to select/

Criterion	Multi-item scale	Single-item scale
Construct concreteness	Abstract	Concrete
Construct dimensionality/complexity	Multidimensional, moderately complex	Unidimensional or extremely complex
Semantic redundancy	Low	High
Primary role of construct	Dependent or independent variable	Moderator or control variable
Desired precision	High	Low
Monitoring changes	Appropriate	Problematic
Sampled population	Homogenous	Diverse
Sample size	Large	Limited

Table 1: Criteria for Assessing the Potential Use of Single-Item Measures

develop a single-item measure with acceptable psychometric properties. In this context, a »difficulty facing the use of single-item scales is the construction of the single-item measures to adequately represent the desired construct. These measures must be constructed carefully to achieve the desired reliability and to complain comparable variance in the dependent variable« (Kwon and Trail, 2005, p. 84). Unfortunately, measurement theory has only paid scant attention to systematic methods for developing and validating single-item measures. This is a gap in need of future research and for which strategies of scale reduction (e.g., see Stanton et al., 2002) or (parts of) Rossiter's (2002) C-OAR-SE scale development procedure may represent useful starting points; the use of item response theory (e.g., see Hambleton and Swaminathan, 1985) is another option worth pursuing as recently suggested by Salzberger (2007).

Relating to the above point, there is a need to document empirical findings on the psychometric qualities of single-item measures in scale handbooks (e.g., see Bearden and Netemeyer, 1999) which should facilitate an increased diffusion and higher acceptance of single-item measures, and would represent a strong argument in favor of the latter in the review process of publications. Furthermore, we suggest that authors developing multi-item scales should, if appropriate, also test and report single-item versions of the latter. This would enable researchers, depending on the detail of the analyses demanded, to choose either a single-item or a full-length version, and give investigators additional confidence in the psychometric soundness of single-item measures.

Finally, it is worth recalling that current measurement practice in management is largely »borrowed from ability-test theory in psychology, where the items differ in *difficulty* and there is within-person variation in *ability* to answer them ... For ability tests a single item cannot provide a precise (reliable) estimate of the individual's ability« (Rossiter, 2002, p. 321, original emphasis). Thus the question arises if we really need highly sophisticated multi-item measures for *all* business-related constructs and treat the latter as if they were psychological (personality and behavioral) and therefore potentially much more complex constructs.[25] Basically, there is no doubt that complex («soft«) constructs in psychology like »depression« or »anxiety« require the need of sound multi-item measurement, but is this also the case with constructs like »purchase likelihood« or »overall satisfaction« with a brand? For the latter, it is entirely conceivable that »theoretical tests and empirical findings would be unchanged if good single-item measures were substituted for these constructs in place of commonly used multiple-item measures« (Bergkvist and Rossiter, 2007, p. 183).

In conclusion, it may well be time to challenge conventional measurement wisdom in management research concerning the (non-) use of single-item measures. While the latter do not, in any way, present a sound measurement option under all circumstances, there *are* conditions under which they do and, therefore, »the use of single-item measures should not be considered fatal flaws in the review process« (Wanous, Reichers and Hudy, 1997, p. 270).

References

Albers, Sönke/Hildebrandt, Lutz (2006): Methodische Probleme bei der Erfolgsfaktorenforschung. Messfehler, formative versus reflektive Indikatoren und die Wahl des Strukturgleichungs-Modells. In: Zeitschrift für betriebswirtschaftliche Forschung (ZfbF), 53. Jg (2006), Februar, S. 2–33.

Anderson, James C./Gerbing, David W. (1982): Some methods for respecifying measurement models to obtain unidimensional construct measurement. In: Journal of Marketing Research, 19. Jg (1982), November, S. 453–460.

Bagozzi, Richard P. (1982): The role of measurement in theory construction and hypothesis testing: Toward a holistic model. In: Fornell C. (Ed.). A second generation of mulitivariate analysis. 1 Jg (1982), New York 1982, S. 5–23.

Bagozzi, Richard P./Heatherton, Todd F. (1994): A general approach to representing multifaceted personality constructs: application to state self-esteem. In: Structural Equation Modeling, 1. Jg (1994), Heft 1, S. 35–67.

Baumgartner, Hans/Homburg, Christian (1996): Applications of structural equation modeling in marketing research: A review. In: International Journal of Research in Marketing, 13. Jg (1996), Heft 2, S. 139–161.

Bearden, William O./Netemeyer, Richard G. (1999): Handbook of marketing scales. 2. Aufl. London 1999.

Bergkvist Lars/Rossiter John R. (2007): The predictive validity of multiple-item versus single-item measures of the same constructs. In: Journal of Marketing Research, 44. Jg (2007), Heft 2, S. 175–184.

Bollen, Kenneth A. (1989): Structural Equations with Latent Variables. New York 1989.

25 Generally, psychological assessment about a person or mood is much more complex and also yields more »effects« than for instance attitudinal measures, all which are closely related. Hence, the complete transfer of psychological measurement procedures should not be fully transferred to measurement in business strategy, since these are two distinct domains.

Bollen, Kenneth A./Lennox, Richard (1991): Conventional wisdom on measurement: A structural equation perspective. In: Psychological Bulletin, 110. Jg (1991), Heft 2, S. 305–314.

Boyd, Brain K./Gove, Steve/Hitt, Michael A. (2005): Construct measurement in strategic management research: Illusion or reality? In: Strategic Management Journal, 26. Jg (2005), Heft 3, S. 239–257.

Boyle, Gerald J. (1991): Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? In: Personality and Individual Differences, 12. Jg (1991), S. 291–294.

Brace, Ian (2004): Questionnaire design: how to plan, structure and write survey material for effective market research. London 2004.

Bryman, Alan/Bell, Emma (2003): Business research methods. New York 2003.

Churchill, Gilbert. A. (1979): A paradigm for developing better measures of marketing constructs. In: Journal of Marketing Research, 16. Jg (1979), S. 64–73.

Cochran, William G. (1977): Sampling techniques, 3. Aufl. New York 1977.

Comrey, Andrew L. (1988): Factor-analytic methods of scale development in personality and clinical psychology. In: Journal of Consulting and Clinical Psychology, 56. Jg (1988), Heft 5, S. 754–761.

Cote, Joseph A./Buckley, Ronald M. (1987): Estimating trait, method, and error variance: generalizing across 70 construct validation studies. In: Journal of Marketing Research, 24. Jg (1987), Heft 3, S. 315–318.

Danes, Jeffrey E./Mann, Karl O. (1984): Unidimensional measurement and structural equation models with latent variables, In: Journal of Business Research, 12. Jg (1984). S. 337–352.

de Boer, A. G. E. M./van Lanschot, J. J. B./Stalmeier, P. F. M./van Sandick, J. W./Hulscher, J. B. F./de Haes J. C. J. M./Sprangers, M. A. G. (2004): Is a single-item visual analogue scale as valid, reliable, and responsive as multi-item scales in measuring quality of life? In: Quality of Life Research, 13. Jg (2004), Heft 2, S. 311–320.

DeVellis, Robert F. (2003): Scale development: theory and application. Applied social research methods series, 2. Aufl., 26 Jg (2003), London 2003.

DeSalvo, Karen B./Fisher, William P./Tran, Ky/Bloser, Nicole/Merrill, William/Peabody, John (2006): Assessing measurement properties of two single-item general health measures. In: Quality of Life Research, 15. Jg (2006), Heft 2, S. 191–201.

Diamantopoulos, Adamantios (2006): The error term in formative measurement models: Interpretation and modeling implications. In: Journal of Modelling in Management, 1. Jg (2006), Heft 1, S. 7–17.

Diamantopoulos, Adamantios/Siguaw, Judy A.(2006): Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. In: British Journal of Management, 17. Jg (2006), S. 263–282.

Diamantopoulos, Adamantios/Riefler, Petra/Roth, Katharina P. (2008): Advancing formative measurement models. In: Journal of Business Research, 58. Jg (2008).

Diamantopoulos, Adamantios/Winklhofer, Heidi M. (2001): Index construction with formative indicators: An alternative to scale development. In: Journal of Marketing Research, 38. Jg (2001), Mai, S. 269–277.

Dillmann, Don A./Sinclair, Michael D./Clark Jon R. (1993): Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail survey. In: Public Opinion Quarterly, 57. Jg (1993), Heft 3, S. 289–304.

Drolet, Aimee L./Morrison, Donald G. (2001a): A practitioner's comment on Aimee L. Drolet and Donald G. Morrison's »Do we really need multiple-item measures in service research?«. In: Journal of Service Research, 3. Jg (2001), Heft 3, S. 196–204.

Drolet, Aimee L./Morrison, Donald G. (2001b): Rejoinder to Grapentine. In: Journal of Service Research, 4. Jg (2001), Heft 2, S. 155–160.

Dolbier, Christyn L./Webster, Judith A./McCalister, Kathrine T./Mallon, Mark W./Steinhardt, Mary A. (2005): Reliability and validity of a single-item measure of job satisfaction. In: American Journal of Health Promotion, 19. Jg (2005), Heft 3, S. 194–198.

Duhachek, Adam/Couglan, Anne T./Iacobucci, Dawn (2005): Results of the standard error of the coefficient alpha index or reliability. In: Marketing Science, 24. Jg (2005), Heft 2, S. 294–301.

Epstein, Seymour (1979): The stability of behavior: I. on predicting most of the people much of the time. In: Journal of Personality and Social Psychology, 37. Jg (1979), July, S. 1097–1126.

Gardner, Donald G./Cummings, Larry L./Dunham, Randall B./Pierce, Jon L. (1998): Single-item versus multiple-item measurement sales: An empirical comparison. In: Educational and Psychological Measurement, 58. Jg (1998), Heft 6, S. 898–915.

Gerbing, David W./Anderson, James C. (1988): An updated paradigm for scale development incorporating unidimensionality and its assessment, In: Journal of Marketing Research, 25. Jg (1988), S. 186–192.

Ginns, Paul/Barrie, Simon (2004): Reliability of single-item ratings of quality in higher education: A replication. In: Psychological Reports, 95. Jg (2004), S. 1023–1030.

Gorsuch, Richard L./McPherson, Susan E. (1989): Intrinsic/extrinsic measurement: i/e-revised single-item scales. In: Journal for the Scientific Study of Religion, 28. Jg (1989), Heft 3, S. 348–354.

Grapentine, Terry H. (2001): A practitioner's comment on Aimee L. Drolet and Doland G. Morrison's: Do we really need multiple-item measures in service research? In: Journal of Service Research, 4. Jg (2001), Heft 2, S. 155–158.

Hambleton, Ronald K. /Swaminathan, Hariharan (1985): Item response theory: Principles and applications. Boston 1985.

Hinkin, Timothy R. (1995): A review of scale development practices in the study of organizations. In: Journal of Management, 21. Jg (1995), Heft 5, S. 967–988.

Homburg, Christian/Giering Annette (1996): Konzeptualisierung und Operationalisierung komplexer Konstrukte – Ein Leitfaden für die Marketingforschung. In: Marketing – Zeitschrift für Forschung und Praxis, 24. Jg (1996), Heft 4, S. 123–137.

Jöreskog, Karl G. (1971): Simultaneous factor analysis in several populations. In: Psychometrika, 36. Jg (1971), Dezember, S. 409–426.

Jöreskog, Karl G./Sörbom Dag (1989): LISREL 7: A guide to the program and applications. Chicago 1989.

Jordan, Jeremy S./Turner, Brian (2005): The development of single-item scales to measure organizational justice and organizational commitment: a reliability and validity assessment. In: Research Consortium Leisure and Recreation and Sport Management, AAHPERD National Convention and Exposition. Chicago 2005.

Krosnick, Jon A. (1999): Survey research. In: Annual Review of Psychology, 50. Jg (1999), S. 537–567.

Kwon, Hyungil/Trail, Galen T. (2005): The feasibility of single-item measures in sport loyalty research. In: Sport Management Review, 8. Jg (2005), S. 69–89.

Law, Kenneth/Wong, Chi-Sum (1999): Multidimensional con-

structs in structural equation analysis: an illustration using the job perception and job satisfaction constructs. In: Journal of Management, 25. Jg (1999), Heft 2, S. 143–160.

Lee, Hanjoon/Delene, Linda M./Bunda, Mary A./Chankon, Kim (2000): Methods of measuring health-care service quality. In: Journal of Business Research, 48. Jg (2000), S. 233–246.

Loo, Robert (2002): A caveat on using single-item versus multiple-item scales. In: Journal of Managerial Psychology, 17. Jg (2002), Heft 1/2, S. 68–75.

Lowenthal, Kate M. (2001): An introduction to psychological tests and scales. 2. Aufl. Hove 2001.

MacKenzie Scott B./Podsakoff Philip M./Jarvis Cheryl B. (2005): The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. In: Journal of Applied Psychology, 90. Jg (2005), Heft 4, S. 710–730.

Moore, Kirstin A./Halle, Tamara G./Vandivere, Sharon/Mariner, Carrie L. (2002): Scaling bach survey scales: How short is too short? In: Sociological Methods & Research, 30. Jg (2002), Heft 4, S. 530–567.

Nagy, Mark S. (2002): Using a single-item approach to measure facet job satisfaction. In: Journal of Occupational and Organizational Psychology, 75. Jg (2002), S. 77–86.

Netemeyer, Richard G./Bearden, William O./Sharma, Subhash (2003): Scaling procedures: Issues and applications. London 2003.

Nunally Jum C. (1967): Psychometric theory. New York 1967.

Nunally Jum C. (1978): Psychometric theory. 2. Aufl. New York 1978.

Nunnally, Jum C. and Bernstein, Ira H. (1994): Psychometric theory. 3. Aufl. New York 1994.

Oshagbemi, Titus (1999): Overall job satisfaction: How good are single versus multiple item measures. In: Journal of Managerial Psychology, 14. Jg (1999), Heft 5, S. 388–403.

Peter, Paul, J. (1979): Reliability: A review of psychometric basics and recent marketing practices. In: Journal of Marketing Research, 16. Jg (1979), Februar, S. 6–17.

Peter, Paul. J. (1981): Construct validity: A review of basic issues and marketing practices. In: Journal of Marketing Research, 18. Jg (1981), Heft 2, S. 133–145.

Podsakoff, Philip M./MacKenzie, Scott B./Podsakoff, Nathan P./Lee, Jeong-Yeon (2003): Common method biases in behavioural research: a critical review of the literature and recommended remedies. Journal of Applied Psychology, 88. Jg (2003), Heft 5, S. 879–903.

Pomeroy, Ian M./Clark, Christopher R./Philip, Ian (2001): The effectiveness of very short scales for depressions screening in elderly medical patients. In: International Journal of Geriatric Psychiatry, 16. Jg (2001), Heft 3, S. 321–326.

Poon, Wai-Yin/Leung, Kwok/Lee, Sik-Yum (2002): The comparison of single item constructs by relative mean and relative variance. In: Organizational Research Methods, 5. Jg (2002), Heft 3, S. 275–298.

Robins, Richard W./Hendin, Holly M./Trzesniewski, Kali H. (2001): Measuring global self-esteem: construct validation of a single-item measure and the Rosenberg self-esteem scale. In: Personality and Social Psychology Bulletin, 27. Jg (2001), Heft 2, S. 151–161.

Rogelberg, Steven G./Luong, Alexandra (1998): Nonresponse to mailed surveys: a review and guide. In: Current Directions in Psychological Science, 7. Jg (1998), Heft 2, S. 60–65.

Rossiter, John R. (2002): The C-OAR-SE procedure for scale development in marketing. In: International Journal of Research in Marketing, 19. Jg (2002), S. 305–335.

Russell, Steven S./Spitzmüller, Christiane/Lin, Lilly F./Stanton, Jeffrey M./Smith, Patricia C./Ironson, Gail H. (2004): Shorter can also be better: The abridged job in general scale. In: Educational and Psychological Measurement, 64. Jg (2004), Heft 5, S. 878–893.

Rust, John/Golombok, Susan (1989): Modern psychometrics. New York 1989.

Sackett, P. R./Larson, James R. Jr. (1990): Research strategies and tactics in industrial and organizational psychology. In Dunnette, M. D./Hough, L. M. (eds.): Handbook of industrial and organization psychology; 2. Aufl. 1. Jg (1990), S. 419–489. Palo Alto 1990.

Salzberger, Thomas (2007): Single-item measures – time for a revival. In: Proceedings of the ANZMAC 2007 Conference, Dunedin, New Zealand 2007.

Scarpello, Vida/Campbell, John P. (1983). Job satisfaction: Are all the parts there? In: Personal Psychology, 36. Jg (1983), S. 577–600.

Shamir, Boas/Kark, Ronit (2004). A single-item graphic scale for the measurement of organizational identification. In: Journal of Occupational and Organizational Psychology, 77. Jg (2004), S. 115–123.

Singh, Jagdip (2003): A reviewer`s gold. In: Journal of the Academy of Marketing Science, 31. Jg (2003), Heft 3, S. 331–336.

Sloan, Jeff A./Aaronson, Neil/Capelleri, Joseph C./Fairclough, Diane L./Varricchio, Claudette (2002): Assessing the clinical significance of single items relative to summated scores. In: Mayo Clinic Proceedings, 77. Jg (2002), Heft 5, S. 497–487.

Smith, Gregory T./McCarthy, Denis M. (1995): Methodological considerations in the refinement of clinical assessment instruments. In: Psychological Assessment, 7. Jg (1995), Heft 3, S. 300–308.

Spector, Paul E. (1992): Summated rating scale construction. Newbury Park, CA 1992.

Stanton, Jeffrey M./Sinar, Evan F./Balzer, William K./Julian, Amada L./Thoresen, Paul/Aziz, Shahnaz/Fisher, Gwenith G./Smith, Patricia C. (2001): Development of a compact measure of job satisfaction: The abridged job descriptive index. In: Educational and Psychological Measurement, 61. Jg (2001), Heft 6, S. 1104–1122.

Stanton, Jeffrey M./Sinar, Evan F./Balzer, William K./Smith Patricia, C. (2002): Issues and strategies for reducing the length of self-report scales. In: Personnel Psychology, 55. Jg (2002), Heft 1, S. 167–193.

Steenkamp, Jan-Benedict E.M./van Trijp, Hans C.M. (1991): The use of LISREL in validating marketing constructs, In: International Journal of Research in Marketing, 8. Jg (1991), S. 283–299.

Steyer, Rolf/Ferring, Dieter/Schmitt, Manfred (1992): States and traits in psychological assessment, In: European Journal of Psychological Assessment, 2. Jg (1992), S. 79–98.

Steyer, Rolf/Schmitt, Manfred (1990): Lantent state-trait models in attitude research, In: Quality and Qauntity, 24. Jg (1990), S. 427–445.

Steyer, Rolf/Schmitt, Manfred/Eid, Michael (1999): Latent state: Trait theory and research in personality and individual differences, In: European Journal of Personality, 13. Jg (1999), S. 389–408.

Sudman, Seymour (1976): Applied sampling. New York 1976.

Temme, Dirk (2006): Die Spezifikation und Identifikation formativer Messmodelle der Marketingforschung in Kovarianzstrukturanalysen, In: ZfP, 28. Jg (2006), Heft 3, S. 183–196.

Van Ryzin, Gregg (2004): The measurement of overall citizen satisfaction. In: Public Performance and Management Review, 27. Jg (2004), Heft 3, S. 9–28.

Viswanathan, Madhu (2005): Measurement error and research design. Sage Publications. London 2005.

Wanous, John. P./Hudy, Michael J. (2001): Single-item reliability: A replication and extension. In: Organizational Research Methods, 4. Jg (2001), Heft 4, S. 361–375.

Wanous, John P./Reichers, Arnon E. (1996): Estimating the reliability of a single-item measure. In: Psychological Reports, 78. Jg (1996), S. 631–645.

Wanous, John P./Reichers, Arnon E./Hudy, Michael J. (1997): Overall job satisfaction: How good are single-item measures? In: Journal of Applied Psychology, 82. Jg (1997), Heft 2, S. 247–252.

Williams, Larry J./Edwards, Jeffrey R./Vandenberg, Robert J. (2003): Recent advances in causal modeling methods for organizational and management research. In: Journal of Management, 29. Jg (2003), Heft 6, S. 903–936.

Wirtz, Jochen/Lee, Meng Chung (2003): An examination of the quality of context-specific applicability of commonly used customer satisfaction measures. In: Journal of Service Research, 5. Jg (2003), Heft 4, S. 345–355.

Youngblut, J. M./Casper, C. R. (1993): Single-item indicators in nursing research. In: Research in Nursing and Health, 16. Jg (1993), Heft 6, S. 459–465.