ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-201902.00024

Single-Item Measurement: Challenges, Responses, and Recommendations

Authors: Wei Xuhua, Zhang Lianghua, Zhang Lianghua

Date: 2019-02-24T00:00:00+00:00

Abstract

Researchers have long debated the use of single-item measures. Proponents argue that single-item measures offer advantages in terms of time and efficiency, while opponents contend that their reliability and validity cannot be guaranteed. Through qualitative and quantitative reviews, this paper summarizes the advantages and disadvantages of single-item measures, analyzes various criticisms from previous research, and responds to them systematically. The systematic review reveals that single-item measures demonstrate acceptable levels of reliability and validity, and that the criterion-related validity of multi-item measures is not significantly superior to that of single-item measures. Finally, it identifies considerations that should be noted during the development and application of single-item measures. Although multi-item measures remain the mainstream approach in current research, future researchers should adopt a more objective perspective toward single-item measures. The academic community should fully understand the potential advantages and appropriate scope of application of single-item measures, thereby enabling them to fulfill their proper role in management psychology and social science research.

Full Text

Single-Item Measures: Queries, Responses, and Recommendations

WEI Xuhua; ZHANG Lianghua
(School of Management, Lanzhou University, Lanzhou 730000, China)

Abstract: The debate surrounding single-item measures has persisted among researchers for decades. Proponents emphasize their advantages in time efficiency and practical utility, while opponents argue that their reliability and validity cannot be adequately guaranteed. Through comprehensive qualitative and quantitative reviews, this paper summarizes the strengths and weaknesses of

single-item measures, systematically examines previous criticisms, and provides targeted responses to each concern. Our systematic review reveals that single-item measures demonstrate acceptable levels of reliability and validity. Notably, the criterion-related validity of multi-item measures does not significantly outperform that of single-item measures. Finally, we identify key considerations for the development and application of single-item measures. Although multi-item measures remain the dominant approach in contemporary research, scholars should adopt a more objective perspective toward single-item measures. The academic community should fully understand the potential advantages and appropriate applications of single-item measures, thereby enabling them to fulfill their proper role in management psychology and social science research.

Keywords: single-item; multi-item; measure; reliability; validity
Classification Codes: B841; B849

Construct measurement has established a solid methodological foundation for the development of psychology and social sciences. When measuring constructs, researchers typically must choose between single-item and multi-item measurement approaches (Ang & Eisend, 2017; Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012; Fuchs & Diamantopoulos, 2009; Sarstedt, Diamantopoulos, & Salzberger, 2016). Single-item measurement refers to the method of evaluating research constructs through a single indicator, characterized by short response time and low cost (Bergkvist & Rossiter, 2007). Multi-item measurement, by contrast, employs two or more items to assess constructs, often yielding higher precision and accuracy (Diamantopoulos et al., 2012; Nair, Ataseven, Habermann, & Dreyfus, 2016).

While both measurement approaches have been widely applied, their prevalence in practice and theory differs substantially. In applied research, management consultants and practitioners favor single-item measures because they significantly reduce respondent refusal rates and lower data collection and processing costs (Bergkvist & Rossiter, 2007; Hoeppner, Kelly, Urbanoski, & Slaymaker, 2011). However, in academic research, scholars prefer multi-item measures. Studies indicate that most research constructs consist of four to six items (Hinkin, 1998), and with the widespread adoption of structural equation modeling techniques, multi-item measures have gained increasing popularity in academia (Diamantopoulos et al., 2012). To this day, many researchers consider single-item measures incomplete and unreliable, arguing that their reliability and validity cannot be effectively guaranteed (Credé, Harms, Niehorster, & Gaye-Valentine, 2012; Roelen et al., 2014; Roelen et al., 2015).

Given the overwhelming dominance of multi-item measures in academic research, scholars have begun comparing the relative merits of single-item and multi-item measures to test whether traditional criticisms of single-item measures are justified. From sporadic comparisons in the 1950s-1980s (Aiken, 1980; Churchill & Peter, 1984; Gorsuch & Mcfarland, 1972; Mosel, 1953) to qualitative and quantitative reviews in recent decades (Diamantopoulos et al., 2012; Fuchs & Diamantopoulos, 2009; Wanous, Reichers, & Hudy, 1997), accumulating ev-

idence suggests that for well-defined constructs, multi-item measures do not significantly outperform single-item measures.

Since Wanous et al.' s (1997) systematic review of single-item measures for job satisfaction, single-item measures have received increasing attention from academic researchers, with their development procedures and application methods becoming increasingly sophisticated (Fisher, Matthews, & Gibbons, 2016; Fuchs & Diamantopoulos, 2009). However, the use of single-item measures in domestic research remains inconsistent, and no studies have specifically discussed single-item measures. This situation hinders Chinese researchers from understanding the latest developments in single-item measures and from using them appropriately. Therefore, this paper provides a systematic review of the development and application of single-item measures, outlines their advantages and disadvantages, analyzes and responds to previous criticisms, and concludes with recommendations for their development and use.

1.1 Advantages of Single-Item Measures

Although multi-item measures remain the mainstream approach in management psychology research, single-item measures offer several advantages over multi-item measures during data collection, including saving time and financial resources (Bergkvist & Rossiter, 2007; Franke, Rapp, & Andzulis, 2013; Gogol et al., 2014; Robins, Hendin, & Trzesniewski, 2001), reducing sample bias, and improving data quality (Postmes, Haslam, & Jans, 2013).

1.1.1 Time Efficiency The time-saving advantage of single-item measures is self-evident, as fewer items require less completion time from respondents (Bergkvist & Rossiter, 2007). This temporal efficiency provides greater flexibility for researchers (Konstabel, Lönnqvist, Walkowitz, Konstabel, & Verkasalo, 2012). Researchers often encounter special research questions (e.g., stress, post-operative pain) and populations (e.g., elderly, children, patients, physicians) that present time-related challenges (Fisher et al., 2016; Gogol et al., 2014; Postmes et al., 2013). For instance, when assessing psychological stress in cancer patients, using multi-item measures increases response time, potentially exacerbating patients' psychological burden and causing negative effects. In such cases, single-item measures often achieve more desirable outcomes.

1.1.2 Cost Reduction Researchers typically compensate respondents based on the number of items completed, a common practice in survey research (Wanous et al., 1997). Multi-item measures require greater financial investment in questionnaire printing (Gogol et al., 2014) and higher incentive payments to ensure diligent completion. Additionally, multi-item measures increase the number of invalid questionnaires due to careless responding (Postmes et al., 2013), further raising research costs. Single-item measures effectively avoid these issues, making them more economical.

1.1.3 Reduced Sample Bias By substantially shortening response time, single-item measures effectively reduce sample bias (Postmes et al., 2013). Many individuals are unwilling to spend time completing lengthy questionnaires that appear unrelated to their interests, causing time constraints to push researchers toward selecting respondents with more free time and thereby introducing sample bias. Conversely, questionnaires containing single-item measures are typically shorter, reducing refusal rates (Bergkvist & Rossiter, 2007). The minimal time requirements of single-item measures allow researchers to select more diverse samples, effectively reducing sample bias (Postmes et al., 2013).

1.1.4 Improved Data Collection Quality Single-item measures enhance data collection quality by minimizing respondents' time burden. On one hand, questionnaires with numerous multi-item measures become lengthy, potentially causing respondents to lose sustained concentration and requiring more time to read and comprehend later items (Postmes et al., 2013), leading to boredom and frustration. In such situations, respondents may approach questions carelessly and perfunctorily (Bergkvist, 2015; Bergkvist & Rossiter, 2007; Credé et al., 2012; Gogol et al., 2014; Zimmerman et al., 2006), severely impacting data quality. Conversely, because single-item measures require less time, they generate less respondent boredom, encouraging more careful reading and answering (Linden & Rosenthal, 2016). On the other hand, during data processing, multi-item measures create more workload and increase researchers' error probability (Reysen, Katzarska-Miller, Nesbit, & Pierce, 2013), whereas single-item measures eliminate many tedious tasks, ensuring efficiency and accuracy.

1.2 Disadvantages of Single-Item Measures

Due to their limited information content, the applicability of single-item measures is restricted. Critics argue that single-item measures have significant limitations when evaluating complex and abstract constructs (Bergkvist & Rossiter, 2007; Fuchs & Diamantopoulos, 2009).

1.2.1 Inability to Fully Capture Complex Constructs For multidimensional complex constructs, a single item cannot simultaneously reflect different dimensions of the construct' s 内涵 (Bergkvist & Rossiter, 2007). Single-item measures often overlook nuanced details. For example, when evaluating job satisfaction with one item, different respondents may automatically weight different aspects of job satisfaction in their evaluation (Fuchs & Diamantopoulos, 2009). However, respondents focus on different aspects and weight them differently, preventing researchers from determining which aspect of job satisfaction is more important (Oshagbemi, 1999). For instance, respondents with low job satisfaction may be dissatisfied due to lack of promotion opportunities or due to unreasonable customer complaints. Clearly, single-item measures cannot identify such specific information, whereas multi-item measures address this problem effectively.

1.2.2 Inability to Accurately Measure Abstract Constructs Compared to multi-item measures, single-item measures are only suitable for evaluating sufficiently concrete constructs such as salary satisfaction, happiness, and anxiety (Gogol et al., 2014; Kim & Abraham, 2017; Wanous et al., 1997). Abstract constructs have vague 内涵, and using single-item measures may cause respondents to make incorrect judgments due to excessive ambiguity and comprehension bias (Bergkvist & Rossiter, 2007). Research shows that respondents'understanding of abstract constructs exhibits substantial heterogeneity (Fuchs & Diamantopoulos, 2009). For abstract constructs such as organizational culture, organizational citizenship behavior, and public service motivation, different respondents may have completely different understandings, thereby affecting measurement precision.

In summary, single-item measures can reduce time and financial costs, ensure data quality by reducing respondent boredom, and improve overall research efficiency (Gogol et al., 2014). However, they also have limitations in applicability, being unsuitable for evaluating complex and abstract constructs (Bergkvist & Rossiter, 2007; Fuchs & Diamantopoulos, 2009). Nevertheless, the greatest criticism of single-item measures concerns scholars' belief that their reliability and validity cannot be adequately guaranteed. In academic review processes, using single-item measures is often considered a major error (Wanous et al., 1997). Many scholars are taught during research methodology training to avoid single-item measures whenever possible (De Boer et al., 2004), and these preconceived notions constitute an important reason why single-item measures face such criticism. However, recent empirical explorations of single-item measure quality have found that previous reliability and validity concerns lack empirical support. Instead, considerable research demonstrates that single-item measures possess acceptable levels of reliability and validity. The following sections address scholars'specific concerns regarding the reliability and validity of single-item measures and provide responses.

2 Concerns and Responses Regarding Reliability of Single-Item Measures

Many researchers believe that the reliability of single-item measures cannot be estimated and that these measures may suffer from serious measurement error, leading most researchers to reject them (Gogol et al., 2014; Van Der Linden & Rosenthal, 2016). In fact, the reliability of single-item measures is not unestimable. With deeper research into single-item measures, evidence regarding their reliability has accumulated, and the academic community needs to reconsider this issue.

2.1 Concerns About Reliability

Reliability is a crucial indicator of a scale' s trustworthiness and stability, with Cronbach' s internal consistency coefficient being most commonly used. Its

calculation formula is as follows:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_i^2}{\sigma_T^2}\right)$$

where k is the number of measurement items, σ_i^2 is the variance of item i' s scores, and σ_T^2 is the variance of the total scale score. According to this formula, calculating Cronbach' s requires more than one item, making it impossible to compute for single-item measures. Additionally, some researchers argue that scale reliability is proportional to the number of items (Churchill & Peter, 1984), causing single-item measures to have lower reliability than multi-item measures (Gogol et al., 2014; Kim & Abraham, 2017; Van Der Linden & Rosenthal, 2016; Wanous et al., 1997). During variable measurement, construct score variance comprises two components: true score and random measurement error. Multi-item measures can reduce random measurement error through mutual offsetting across items, whereas single-item measures, having only one item, cannot achieve this effect (Fu, 2005; Konrath, Meier, & Bushman, 2014; Robins et al., 2001). Therefore, single-item measures may suffer from more serious random measurement error, resulting in lower reliability.

2.2 Responses to Reliability Concerns

Although reliability is a common criticism of single-item measures, researchers can evaluate their reliability through alternative indicators such as test-retest reliability or by deriving reliability through formulas (Fisher et al., 2016; Spörrle & Bekk, 2014; Wanous et al., 1997).

2.2.1 Test-Retest Reliability Test-retest reliability reflects a scale' s temporal stability, typically calculated as the correlation coefficient between scores from two time points (Fisher et al., 2016). This method is one of the most widely used approaches for assessing single-item measure reliability. Table 1 lists test-retest reliabilities for 60 single-dimensional constructs. As shown, 68.33% of single-dimensional single-item measures have test-retest reliability above 0.70, with a median of 0.73 and a sample-weighted average of 0.71, indicating that single-item measures are generally reliable.

2.2.2 Backward-Calculated Reliability The reliability of variables in a research sample affects estimates of true relationships between variables. In scientific research, the true correlation between two variables can be obtained through reliability-corrected sample correlations (Postmes et al., 2013). The relationship between true correlation and sample correlation is as follows (Loo, 2002; Postmes et al., 2013; Wanous et al., 1997):

$$\rho = \frac{r_{xy}}{\sqrt{r_{xx} \cdot r_{yy}}}$$

where r_{xx} is variable x' s reliability in the sample, r_{yy} is variable y' s reliability, ρ is the true correlation between x and y, and r_{xy} is the sample correlation between x and y. Due to measurement error, sample correlation r_{xy} is typically smaller than true correlation ρ.

If we assume x and y are two different measures of the same construct, where x is a single-item measure and y is a multi-item measure, then single-item measure reliability can be calculated using the following formula:

$$r_{xx} = \frac{r_{xy}^2}{\rho^2 \cdot r_{yy}}$$

In this formula, since single-item measure x and multi-item measure y assess the same construct, the maximum value of their true correlation ρ is 1 (Postmes et al., 2013; Wanous & Hudy, 2001). This yields the minimum reliability for single-item measures:

$$r_{xx(min)} = \frac{r_{xy}^2}{r_{yy}}$$

Most previous empirical studies on single-item measures have reported the correlation coefficient r_{xy} between single-item and multi-item measures and the reliability r_{yy} of multi-item measures. Therefore, researchers can use this method to derive the minimum reliability of single-item measures (Loo, 2002; Postmes et al., 2013; Wanous & Hudy, 2001; Wanous et al., 1997).

Using this approach, this study statistically analyzed the minimum reliability of single-item measures from previous empirical research[2], with results shown in Table 1 . As shown, only 15.38% of single-dimensional constructs have minimum reliability below 0.60, with the remainder above 0.60, and the median minimum reliability for single-dimensional single-item measures is 0.67. It should be noted that minimum reliability does not represent the actual reliability of single-item measures. In practice, single-item measure reliability is typically higher than this minimum value. For example, based on data from Núñez-Peña, Guilera, and Suárez-Pellicioni (2014), the derived minimum reliability for a single-item math anxiety measure is 0.63, while the reported test-retest reliability in their study is 0.81, substantially higher than the minimum reliability. However, for multidimensional constructs, both test-retest and minimum reliability values are generally lower, with medians of 0.62 and 0.50 respectively, suggesting potential problems when using single-item measures to evaluate multidimensional constructs. Thus, single-item measure reliability is not as unacceptable as previously suggested, particularly when evaluating unidimensional constructs.

3 Concerns and Responses Regarding Validity of Single-Item Measures

Another reason for the unacceptability of single-item measures concerns questions about their validity. Validity is a crucial indicator of a measurement' s effectiveness and accuracy. The primary limitation of single-item measure validity lies in their lack of comprehensiveness (Fisher et al., 2016; Oshagbemi, 1999). Because single-item measures cannot provide sufficiently diverse and detailed responses, some researchers argue that they overlook subtle differences among participants (Bergkvist & Rossiter, 2007; Diamantopoulos et al., 2012; Konrath et al., 2014; Kwon & Trail, 2005), causing measurement bias and reducing explanatory power. Critics have raised concerns about content validity, construct validity, and criterion-related validity.

3.1 Concerns About Validity

3.1.1 Content Validity Concerns Content validity refers to the degree to which a measurement tool reflects the 内涵 and scope of the target construct. From a content validity perspective, single-item measures are typically used for unidimensional constructs. Although not comprehensive, they can still capture the core 内涵 of unidimensional constructs (Robins et al., 2001). For example, researchers often use single-item measures to evaluate constructs such as satisfaction, quality of life, and well-being (Cheung & Lucas, 2014; De Boer et al., 2004; Zimmerman et al., 2006). However, for multidimensional constructs, researchers typically use multiple items to evaluate various dimensions to measure all aspects as completely as possible. Obviously, a single item' s incompleteness and one-sidedness become particularly pronounced when reflecting multidimensional constructs, substantially reducing content validity.

3.1.2 Construct Validity Concerns Construct validity examines whether a scale accurately measures the target construct, typically reflected through convergent and discriminant validity (Fisher et al., 2016). Researchers usually evaluate convergent and discriminant validity using composite reliability and average variance extracted. Since single-item measures cannot calculate composite reliability and average variance extracted (Wohlgemuth & Wenzel, 2016), many researchers believe that convergent and discriminant validity cannot be assessed for single-item measures (Fuchs & Diamantopoulos, 2009).

3.1.3 Criterion-Related Validity Concerns Regarding criterion-related validity, single-item measures may produce weaker explanatory power for criterion variables. Some scholars note that single-item measures contain very limited information (Fisher et al., 2016; Oshagbemi, 1999), preventing comprehensive and accurate measurement of unidimensional or multidimensional constructs and thereby affecting their predictive utility. Multi-item measures typically contain more items and information, better reflecting the nature of

research constructs and thus demonstrating better predictive validity (Gogol et al., 2014; Van Der Linden & Rosenthal, 2016).

For these reasons, many scholars express concerns about the validity of single-item measures and prefer multi-item measures.

3.2 Responses to Validity Concerns

Although researchers have raised some validity concerns about single-item measures, these criticisms largely remain at the conceptual level. In fact, numerous empirical studies in recent years have found that the validity of single-item measures is acceptable (Elo, Leppänen, & Jahkola, 2003; Konrath et al., 2014; Konstabel et al., 2017; Kwon & Trail, 2005; Robins et al., 2001).

3.2.1 Content Validity Responses While multi-item measures may be more comprehensive, research finds that additional items contribute minimal extra information (Fuchs & Diamantopoulos, 2009). Multi-item measures typically include multiple similar yet slightly different items to enhance comprehensiveness. In practice, this similarity often causes respondent boredom, leading participants to select identical responses throughout a questionnaire (Nair et al., 2016; Robins et al., 2001), reducing the likelihood of obtaining additional information. Furthermore, similar items pose another problem: if these items are biased, repeated presentation doubles this bias, further reducing multi-item measure accuracy.

Recent empirical studies have confirmed that single-item measures demonstrate acceptable content validity when evaluating some constructs (Elo et al., 2003; Fisher et al., 2016; Jordan & Turner, 2008; Robins et al., 2001). For example, Fisher et al. (2016) used expert rating methods to evaluate single-item measure content validity and found their newly developed single-item measures had good content validity. Even for multidimensional constructs, if multiple dimensions can form an overall variable, single-item measures can have sufficient content validity (Robins et al., 2001). Additionally, researchers have increased content validity by using descriptions containing multiple adjectives. For instance, Woods and Hampson (2005) used the following reverse-scored single-item measure to assess the extraversion dimension of the Big Five personality traits: "xxx is a reserved, introverted person who dislikes attracting attention and is relatively shy among strangers." Their results also found little difference between this scale and multi-item measures.

3.2.2 Construct Validity Responses To address concerns about the inability to evaluate single-item measure construct validity, researchers have developed new assessment methods. First, building on the multitrait-multimethod (MTMM) matrix, researchers calculate correlation coefficients between single-item measures and corresponding multi-item measures to assess convergent validity (Fisher et al., 2016; Postmes et al., 2013), reflecting the extent to which single-item measures can replace multi-item measures. High correlation between

single-item and multi-item measures indicates strong convergent validity. Second, researchers evaluate discriminant validity by calculating correlations between single-item measures and similar constructs (Fisher et al., 2016; Nichols & Webster, 2013; Spörrle & Bekk, 2014). Low correlations with these similar constructs indicate good discriminant validity.

To better demonstrate single-item measure construct validity, this study quantitatively integrated convergent and discriminant validity from previous empirical research through meta-analysis[3]. Table 2 presents quantitative analysis results for single-item measure convergent and discriminant validity. Convergent validity represents the correlation between single-item and multi-item measures, while discriminant validity represents correlations between single-item measures and similar constructs. As shown, single-item measures correlate highly with multi-item measures (=0.72, p<0.001), demonstrating good convergent validity. Furthermore, correlations between single-item measures and similar constructs are low (=0.22, p<0.001), indicating good discriminant validity. Therefore, previous concerns about single-item measure construct validity are not supported.

3.2.3 Criterion-Related Validity Responses Even in multi-item measures, individual items may vary in quality. During construct assessment, good items can mask poor ones (Bergkvist & Rossiter, 2007; Fisher et al., 2016; Van Der Linden & Rosenthal, 2016). However, single-item measures are typically carefully selected by experts or researchers, avoiding unnecessary noise and ensuring predictive validity. Additionally, many researchers have compared criterion-related validity differences between single-item and corresponding multi-item measures to evaluate their relative merits. Numerous empirical studies have found that single-item measures have similar criterion-related validity to corresponding multi-item measures (Ang & Eisend, 2017; Gogol et al., 2014; Kwon & Trail, 2005; Riordan et al., 2018; Robins et al., 2001; Woods & Hampson, 2005). To better demonstrate this finding, this study used meta-analysis to quantitatively compare criterion-related validity between single-item and multi-item measures (see Table 3). As shown, the relationship between single-item measures and criterion variables (=0.12, p<0.001) does not differ significantly from that between multi-item measures and criterion variables (=0.16, p<0.001) (QB=0.92, p=0.34). Therefore, from a criterion-related validity perspective, multi-item measures do not significantly outperform single-item measures.

4 Recommendations for Developing and Using Single-Item Measures

The preceding review demonstrates that although single-item measures have certain limitations, sufficient evidence supports their reliability and validity. Therefore, when literature support and empirical evidence are adequate, researchers may select single-item measures for their studies.

4.1 Development Recommendations for Single-Item Measures

Researchers can either select an item from existing multi-item measures to evaluate constructs (Diamantopoulos et al., 2012; Fisher et al., 2016; Fuchs & Diamantopoulos, 2009; Loo, 2002) or develop a new item for construct evaluation (Cheung & Lucas, 2014; Gogol et al., 2014; Robins et al., 2001; Van Der Linden & Rosenthal, 2016).

4.1.1 Selecting Single-Item Measures from Existing Scales Researchers can select the item with the highest factor loading from existing multi-item measures with similar or semantically overlapping content as the single-item measure indicator (Diamantopoulos et al., 2012). This method is relatively objective and not subject to subjective interference. However, since the item with the highest factor loading may vary across samples, researchers can select the optimal item across samples using sample-weighted average loadings. Additionally, researchers can use expert rating methods to select the item with highest content validity from existing multi-item measures (Aiken, 1980; Sarstedt et al., 2016; Sarstedt, Diamantopoulos, Salzberger, Baumgartner, & Woodside, 2016). Researchers can invite multiple experts to rate the extent to which each item in a multi-item measure reflects the construct' s core 内涵. Subsequently, researchers can calculate inter-rater consistency using the within-group agreement index (rwg). Finally, researchers can select the item with rwg above 0.7 and the highest average expert rating as the final single-item measure.

4.1.2 Developing New Single-Item Measures Researchers can also develop new single-item measures to evaluate constructs. Although items extracted from multi-item measures may correlate more highly with original multi-item measures and show greater similarity in criterion-related validity, their descriptions are often less comprehensive than newly developed single-item measures (Fisher et al., 2016). When generating new single-item measures, researchers must first have a correct and clear understanding of the target construct to determine whether it is suitable for single-item measurement. Researchers should recognize that a good single-item measure should: accurately reflect the construct' s core 内涵, be applicable across different samples, and effectively discriminate from other constructs (Diamantopoulos et al., 2012). Therefore, when describing single-item measures, researchers must carefully consider wording based on construct meaning, particularly when measuring academic constructs unfamiliar to the general public. Otherwise, the generality of single-item measure descriptions and the specialized nature of constructs may prevent respondents from understanding what researchers intend to measure, requiring additional written or oral explanations that waste time (Credé et al., 2012).

It should be noted that current empirical evidence has not found significant effects of single-item measure response scales on measurement results (Kim & Abraham, 2017; Robins et al., 2001), so researchers need not worry excessively about scale effects.

4.2 Usage Recommendations for Single-Item Measures

Researchers should not blindly resist single-item measures but should recognize their potential advantages. Simultaneously, researchers should clearly understand that single-item measures are not suitable for every situation.

4.2.1 Single-Item Measures Are Better for Unidimensional Constructs Compared to multidimensional constructs, single-item measures are more suitable for unidimensional constructs. In much management psychology research, constructs typically contain multiple non-integrable dimensions, such as personality and conflict management styles. In these cases, single-item measures contain very limited information and cannot comprehensively reflect all dimensions of multidimensional constructs through one item. However, if multiple dimensions of a multidimensional construct can form an overall variable (e.g., global self-esteem), single-item measures can be used (Robins et al., 2001).

4.2.2 Single-Item Measures Are Better for Concrete and Clear Constructs Single-item measures are more appropriate for sufficiently concrete constructs than for overly broad or vague ones. Some research constructs in social sciences, such as "mindfulness,"are relatively abstract and poorly understood by the general public, making them unsuitable for single-item measurement. Although single-item measures are not ideal for evaluating multidimensional constructs, if each dimension of a multidimensional construct is sufficiently clear, researchers can use single-item measures to evaluate each dimension separately. For example, researchers frequently use single-item measures to assess each personality trait in the Big Five (Konstabel et al., 2017; Konstabel et al., 2012; Woods & Hampson, 2005) and to evaluate attitudinal constructs (e.g., job satisfaction, life satisfaction) (Cheung & Lucas, 2014; De Boer et al., 2004; Zimmerman et al., 2006) and emotional constructs (joy, anger, sorrow, happiness) (Fisher & To, 2012; Riordan et al., 2018).

4.2.3 Single-Item Measures Are Better for Time-Constrained Research When researchers face resource and time constraints, they should consider using single-item measures to assess relevant constructs. Single-item measures not only effectively save time but also address issues arising from time constraints, including sample bias, resource waste, and careless responding. Single-item measures are particularly suitable for time-sensitive experimental research, longitudinal studies, and experience sampling method studies (Konstabel et al., 2012; Robins et al., 2001). In many experimental studies, to ensure manipulation effectiveness and minimize interference from irrelevant factors, researchers adopt single-item measures to reduce experimental duration. For longitudinal and experience sampling method studies, where researchers repeatedly collect data at specific time points over a period (Fisher & To, 2012), single-item measures significantly reduce respondent frustration and perfunctory responding (Fisher et al., 2016).

4.2.4 Single-Item Measures Are Better for Large-Scale Studies Many mainstream large-scale cross-national tracking surveys (e.g., World Values Survey) involve numerous respondents across many countries. Researchers aim to collect as much data as possible, resulting in surveys covering many constructs. If all such surveys used multi-item measures, they would impose tremendous burden on respondents. For large-scale survey research, reducing participant burden is a primary consideration (Cheung & Lucas, 2014), and single-item measures address this issue effectively. Single-item measures enable researchers to cover as many constructs as possible while minimizing respondent burden. Therefore, single-item measures are commonly used construct evaluation methods in large-scale studies.

4.2.5 Single-Item Measures Are Better for Non-Core Constructs in Research If a construct is not the researcher' s primary focus or does not require high precision, such as when used as a control variable, researchers should prioritize single-item measures (Fisher et al., 2016; Fuchs & Diamantopoulos, 2009). This approach allows researchers to include as many control variables as possible and better explain the unique contributions of their research variables based on existing studies.

5 Conclusion

Through qualitative and quantitative reviews, this study summarized the advantages and disadvantages of single-item measures and responded to various academic concerns about their reliability and validity. The findings indicate that single-item measures have acceptable levels of reliability and validity, and that multi-item measures do not significantly outperform single-item measures in criterion-related validity. Although multi-item measures remain the mainstream approach in current research, future researchers should adopt a more objective perspective toward single-item measures. Researchers should recognize the potential advantages and appropriate applications of single-item measures, enabling them to fulfill their proper role in management psychology and social science research.

Acknowledgments: We thank Associate Professor Kelly Z. Peng from the Department of Business Administration at Hong Kong Shue Yan University for revising the English abstract of this paper!

*Indicates literature included in quantitative analysis

References

*Abdelkhalek, A. M. (2006). Measuring happiness with a single-item scale. Social Behavior and Personality: An International Journal, 34(2), 139-150.

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. Educational and Psychological Measurement, 40(4), 955-959.

Ang, L., & Eisend, M. (2017). Single versus multiple measurement of attitudes: A meta-analysis of advertising studies validates the single-item measure approach. Journal of Advertising Research, 58(3), 1–10.

Bergkvist, L. (2015). Appropriate use of single-item measures is here to stay. Marketing Letters, 26(3), 245–247.

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(2), 175–184.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). Introduction to meta-analysis. United Kingdom: John Wiley & Sons.

*Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. Quality of Life Research, 23(10), 2809–2818.

Churchill, G. A., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. Journal of Marketing Research, 21(4), 360–375.

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the big five personality traits. Journal of Personality and Social Psychology, 102(4), 874–888.

*De Boer, A. G. E. M., Van Lanschot, J. J. B., Stalmeier, P. F. M., Van Sandick, J. W., Hulscher, J. B., De Haes, J. C., et al. (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? Quality of Life Research, 13(2), 311–320.

Denissen, J. J. A., Geenen, R., Selfhout, M., & Aken, M. A. G. V. (2008). Single-item big five ratings in a social network design. European Journal of Personality, 22(1), 37–54.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. Journal of the Academy of Marketing Science, 40(3), 434–449.

Elo, A. L., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. Scandinavian Journal of Work Environment & Health, 29(6), 444–451.

Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. Journal of Organizational Behavior, 33(7), 865–877.

*Fisher, G. G., Matthews, R. A., & Gibbons, A. M. (2016). Developing and investigating the use of single-item measures in organizational research. Journal of Occupational Health Psychology, 21(1), 3–23.

Franke, G. R., Rapp, A., & Andzulis, J. M. (2013). Using shortened scales in sales research: Risks, benefits, and strategies. Journal of Personal Selling & Sales Management, 33(3), 319–328.

Fu, Y. (2005). Measuring personal networks with daily contacts: A single-item survey question and the contact diary. Social Networks, 27(3), 169–186.

Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. Die Betriebswirtschaft, 69(2), 195–210.

*Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: an empirical comparison. Educational and Psychological Measurement, 58(6), 898–915.

*Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., et al. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. Contemporary Educational Psychology, 39(3), 188–205.

Gorsuch, R. L., & Mcfarland, S. G. (1972). Single vs. Multiple-item scales for measuring religious values. Journal for the Scientific Study of Religion, 11(1), 53–64.

*Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. Journal of Research in Personality, 37(6), 504–528.

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? Quality of Life Research, 13(2), 311–320.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. Organizational Research Methods, 1(1), 104–121.

*Hoeppner, B. B., Kelly, J. F., Urbanoski, K. A., & Slaymaker, V. (2011). Comparative utility of a single-item vs. Multiple-item measure of self-efficacy in predicting relapse among young adults. Journal of Substance Abuse Treatment, 41(3), 305–312.

*Jordan, J. S., & Turner, B. A. (2008). The feasibility of single-item measures for organizational justice. Measurement in Physical Education and Exercise Science, 12(4), 237–257.

*Kim, H. J., & Abraham, I. (2017). Measurement of fatigue: Comparison of the reliability and validity of single-item and short measures to a comprehensive measure. International Journal of Nursing Studies, 65, 35–43.

*Konrath, S., Meier, B. P., & Bushman, B. J. (2018). Development and validation of the single item trait empathy scale (SITES). Journal of Research in Personality, 73, 111–122.

*Konrath, S., Meier, B. P., & Bushman, B. J. (2014). Development and validation of the single item narcissism scale (SINS). Plos One, 9(8), e103459.

Konstabel, K., Lönnqvist, J. E., Leikas, S., García, V. R., Qin, H., Verkasalo, M., et al. (2017). Measuring single constructs by single items: Constructing an even shorter version of the "short five" personality inventory. Plos One, 12(8), e0182714.

Konstabel, K., Lönnqvist, J. E., Walkowitz, G., Konstabel, K., & Verkasalo, M. (2012). The 'short five' (S5): Measuring personality traits using comprehensive single items. European Journal of Personality, 26(1), 13–29.

*Kwon, H., & Trail, G. (2005). The feasibility of single-item measures in sport loyalty research. Sport Management Review, 8(1), 69–88.

Linden, S. V. D., & Rosenthal, S. A. (2016). Measuring narcissism with a single question? A replication and extension of the single-item narcissism scale (SINS). Personality and Individual Differences, 90(3), 238–241.

*Loo, R. (2002). A caveat on using single-item versus multiple-item scales. Journal of Managerial Psychology, 17(1), 68–75.

*Macias, C., Gold, P. B., Öngür, D., Cohen, B. M., & Panch, T. (2015). Are single-item global ratings useful for assessing health status? Journal of Clinical Psychology in Medical Settings, 22(4), 313–321.

Matthias, S., & Magdalena, B. (2014). Meta-analytic guidelines for evaluating single-item reliabilities of personality instruments. Assessment, 21(3), 272–285.

Mosel, J. N. (1953). Single-item tests for personnel screening. Educational and Psychological Measurement, 13(2), 179–192.

*Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). The single-item math anxiety scale (SIMA): An alternative way of measuring mathematical anxiety. Personality and Individual Differences, 60(4), S75–S76.

*Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. Journal of Occupational and Organizational Psychology, 75(1), 77–86.

Nair, A., Ataseven, C., Habermann, M., & Dreyfus, D. (2016). Toward a continuum of measurement scales in just-in-time (JIT) research -an examination of the predictive validity of single-item and multiple-item measures. Operations Management Research, 9(1–2), 35–48.

*Nichols, A. L., & Webster, G. D. (2013). The single-item need to belong scale. Personality and Individual Differences, 55(2), 189–192.

Oshagbemi, T. (1999). Overall job satisfaction: How good are single versus multiple-item measures? Journal of Managerial Psychology, 14(5), 388–403.

*Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. British Journal of Social

Psychology, 52(4), 597–617.

Reysen, S., Katzarska-Miller, I., Nesbit, S. M., & Pierce, L. (2013). Further validation of a single-item measure of social identification. European Journal of Social Psychology, 43(6), 463–470.

Riordan, B. C., Cody, L., Flett, J. A. M., Conner, T. S., Hunter, J., & Scarf, D. (2018). The development of a single item FoMO (Fear of Missing Out) scale. Current Psychology, doi: 10.1007/s12144-018-9824-8.

*Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg self-esteem scale. Personality and Social Psychology Bulletin, 27(2), 151–161.

Roelen, C. A., Van, R. W., Groothoff, J. W., Jj, V. D. K., Twisk, J. W., & Heymans, M. W. (2014). Work ability as prognostic risk marker of disability pension: Single-item work ability score versus multi-item work ability index. Scandinavian Journal of Work Environment & Health, 40(4), 428–431.

Roelen, C. A. M., Heymans, M. W., Twisk, J. W. R., Laaksonen, M., Pallesen, S. L., Magerøy, N., et al. (2015). Health measures in prediction models for high sickness absence: Single-item self-rated health versus multi-item SF-12. European Journal of Public Health, 25(4), 694–699.

*Russell, J. A. (1989). Affect grid: a single-item scale of pleasure and arousal. Journal of Personality and Social Psychology, 57(3), 493–502.

Sarstedt, M., Diamantopoulos, A., & Salzberger, T. (2016). Should we use single items? Better not. Journal of Business Research, 69(8), 3199–3203.

Sarstedt, M., Diamantopoulos, A., Salzberger, T., Baumgartner, P., & Woodside, A. G. (2016). Selecting single items to measure doubly concrete constructs: A cautionary tale. Journal of Business Research, 69(8), 3159–3167.

Spörrle, M., & Bekk, M. (2014). Meta-analytic guidelines for evaluating single-item reliabilities of personality instruments. Assessment, 21(3), 272–285.

*Van Der Linden, S., & Rosenthal, S. A. (2016). Measuring narcissism with a single question? A replication and extension of the single-item narcissism scale (SINS). Personality and Individual Differences, 90(3), 238–241.

*Wanous, J. P., & Hudy, M. J. (2001). Single-item reliability: A replication and extension. Organizational Research Methods, 4(4), 361–375.

*Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? Journal of Applied Psychology, 82(2), 247–252.

Wohlgemuth, V., & Wenzel, M. (2016). Dynamic capabilities and routinization. Journal of Business Research, 69(5), 1944–1948.

*Woods, S. A., & Hampson, S. E. (2005). Measuring the big five with single items using a bipolar response scale. European Journal of Personality, 19(5),

373–390.

Zimmerman, M., Ruggero, C. J., Chelminski, I., Young, D., Posternak, M. A., Friedman, M., et al. (2006). Developing brief scales for use in clinical practice: The reliability and validity of single-item self-report measures of depression symptom severity, psychosocial impairment due to depression, and quality of life. Journal of Clinical Psychiatry, 67(10), 1536–1541.

Appendix: Literature Search, Coding, and Analysis Process for Quantitative Analysis

1. Literature Search Process

Researchers searched published literature on single-item measures in Web of Science, Google Scholar, EBSCO, Springer, Sage, Elsevier, CNKI, and Baidu Scholar using keywords "single-item measure" or "single-indicator measure," with publication dates limited to January 1988–June 2018. Additionally, we reviewed single-item measure review articles to ensure inclusion of as many relevant papers as possible. Retrieved articles meeting any of the following criteria were included in our quantitative analysis: (1) reported test-retest reliability of single-item measures; (2) reported correlation between single-item measures and multi-item measures of the same construct; (3) reported correlation between single-item measures and similar constructs; (4) reported correlations of both single-item and multi-item measures with criterion variables. Based on these criteria, we identified 29 quantitative research papers on single-item measures. These papers reported 79 reliability-related values, 45 effect sizes testing convergent validity, 51 effect sizes testing discriminant validity, 63 effect sizes for single-item measure criterion-related validity, and 79 effect sizes for multi-item measure criterion-related validity.

2. Variable Coding Process

Two researchers independently coded effect sizes based on the above criteria, with initial consistency exceeding 90%. Discrepancies were jointly reviewed and corrected by both authors to produce consistent coding results. In reliability coding, test-retest reliabilities of single-item measures were used for reliability statistics. Additionally, if a paper reported both the correlation between single-item and multi-item measures of the same construct and the reliability coefficient of the multi-item measure, we could calculate the backward-derived minimum reliability of the single-item measure (see Section 2.2.2 for formula). In validity coding, if authors used both single-item and multi-item measures to assess the same construct and reported their correlation, we calculated the weighted average effect size for convergent validity. If a paper reported correlations between single-item measures and similar constructs, we calculated the weighted average effect size for discriminant validity. If a paper reported correlations of both single-item and multi-item measures with criterion variables, we calculated

weighted average effect sizes for criterion-related validity and tested differences between single-item and multi-item measures. When multiple similar measures existed for the same construct in the same sample, we combined these effect sizes to reduce artificial inflation of sample size. We also collected multi-item measure reliabilities and corrected effect sizes for measurement error. For multi-item measures without reported reliability, we substituted the average weighted reliability from other multi-item measures (0.762).

3. Analysis Process

This study's reliability analysis of single-item measures involved frequency statistics, with results shown in Table 1 in the main text. Validity analysis employed meta-analysis using CMA2.0 software, with results shown in Tables 2 and 3 in the main text. During meta-analysis, we used fail-safe numbers to assess publication bias severity. Results in Tables 2 and 3 show that fail-safe numbers for each analysis exceeded the number of effect values, indicating that publication bias was not severe. For meta-analysis model selection, following Borenstein, Hedges, Higgins, and Rothstein (2011), we selected random-effects models, which assume that effect sizes vary randomly across studies rather than being fixed, providing better alignment with real-world conditions.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.