

Can't we make it any shorter? The limits of personality assessment and way to overcome them

Rammstedt, Beatrice; Beierlein, Constanze

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Rammstedt, B., & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and way to overcome them. *Journal of Individual Differences*, 35(4), 212-220. <https://doi.org/10.1027/1614-0001/a000141>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

gesis
Leibniz-Institut
für Sozialwissenschaften

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der

Leibniz-Gemeinschaft

Can't We Make It Any Shorter?

The Limits of Personality Assessment and Way to Overcome Them

Beatrice Rammstedt and Constanze Beierlein

GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

Abstract. Psychological constructs are becoming increasingly important in social surveys. Scales for the assessment of these constructs are usually developed primarily for individual assessment and decision-making. Hence, in order to guarantee high levels of reliability, measurement precision, and validity, these scales are in most cases much too long to be applied in surveys. Such settings call for extremely short measures validated for the population as a whole. However, despite the unquestionable demand, appropriate measures are still lacking. There are several reasons for this. In particular, short scales have often been criticized for their potential psychometric shortcomings with regard to reliability and validity. In this article, the authors discuss the advantages of short scales as alternative measures in large-scale surveys. Possible reasons for the assumed limited psychometric qualities of short scales will be highlighted. The authors show that commonly used reliability estimators are not always appropriate for judging the quality of scales with a minimal number of items, and they offer recommendations for alternative estimation methods and suggestions for the construction of a thorough short scale.

Keywords: short scales, personality assessment, psychometric quality, survey research

Over the past years, psychological constructs such as personality, locus of control, basic human values, or intelligence have attracted increasing attention among researchers outside the core field of psychological research. The power of psychological constructs to predict social or socioeconomic outcomes has led these researchers to consider psychological variables in their research. Several empirical studies corroborate the predictability of psychological variables (Gottfredson, 1997; Gottfredson & Deary, 2004; Schmidt & Hunter, 1998; Strenze, 2007). For example, studies have shown that cognitive ability is the best predictor of success in life. Individuals with comparatively higher cognitive ability are more successful in education, at work, and in their private lives. Hence, more intelligent individuals earn, on average, higher incomes, hold more senior positions, are less often unemployed, and have a lesser tendency to divorce and a lower delinquency rate. Numerous other socially and socioeconomically relevant indicators are also significantly related to psychological constructs. For example, political voting behavior can be predicted by psychological characteristics such as personality traits (Mondak, 2010; Schoen & Steinbrecher, 2013). In addition, political (self-)efficacy serves as one of the best predictors when it comes to explaining individual differences in political behavior (Caprara, Vecchione, Capanna, & Mebane, 2009; see also Bandura, 1997). Health - and even mortality risk - has been shown to be linked to such psychological characteristics. More conscientious individuals have been found to have, on average, better health and a lower mortality risk (Allison, Guichard, Fung, & Gilain, 2003; Arthur & Graziano, 1996; Rasmussen, Scheier, & Greenhouse, 2009; for a more comprehensive overview of the relations between psychological constructs and social and socioeconomic outcome variables see Kemper, Beierlein, Kovaleva, & Rammstedt, 2012; Rammstedt, Kemper, & Schupp, 2013). Given the apparent usefulness of psychological characteristics for enhancing the predictability of social and socioeconomic variables, the Nobel laureate in economics James Heckman suggested that social and socioeconomic studies should incorporate validated instruments for the measurement of cognitive and noncognitive skills, such as personality traits (Borghans, Duckworth, Heckman, & Weel, 2008). Several other researchers and institutions draw the same conclusions as the Nobel prize winner Heckman (e.g., German Data Forum, 2010; Goldberg, 2005; Rammstedt, 2010).

Assessing Psychological Constructs in Large-Scale Social Surveys

During the last decades, several attempts have been made to include psychological variables in social and economic surveys. For example, the German Socio-Economic Panel (SOEP) started assessing locus of control as early as the 1990s. In recent years, measures for risk aversion, the Big Five personality dimensions, intelligence, and a number of other major psychological variables have been incorporated into the study (Lohmann, Spieß, Groh-Samberg, & Schupp, 2009; Schupp, Spieß, & Wagner 2008). Other large-scale panel or cross-sectional studies, such as the International Social Survey Programme (ISSP), the European Social Survey (ESS), the German National Education Panel Study (NEPS), the German Longitudinal Election Study (GLES), the German Panel Study "Labour Market and Social Security" (PASS), the Household, Income and Labour Dynamics in Australia (HILDA) Survey, the UK Household Longitudinal Study (UKHLS), and the DNB Household Survey (DHS), have followed this trend and now include psychological variables in their questionnaires (cf. Kemper, Beierlein, Kovaleva, et al., 2012; Rammstedt, Kemper, & Schupp, 2013; for an overview, see Rammstedt & Spinath, 2013). Moreover, the German National Cohort study, which is just

getting underway, will also include psychological measures. Based on approximately 200,000 respondents, it will be the largest and most comprehensive panel study in Germany.

Objectives of the Present Article

As this brief overview shows, there is a clear and growing demand for validated and standardized measures for the assessment of key psychological variables in interdisciplinary survey research. Despite this unquestionable demand, appropriate measures are still lacking. Although there is a large pool of psychometrically tested instruments for nearly all psychological constructs, most are unsuitable for large-scale surveys. Against this background, the present article has three central objectives: (1) First of all, we explain the limited usability of traditional psychological measures for large-scale social surveys and the need for properly developed survey-compatible psychological short scales as potential alternatives. (2) Second, we discuss potential limitations of psychological short scales with a special focus on their psychometric properties (reliability, validity). (3) Finally, we suggest ways to handle or overcome potential psychometric shortcomings of psychological short scales.

The Need for Properly Developed Psychological Short Scales in Large-Scale Social Surveys

The majority of typical psychological instruments has been developed and validated for the diagnosis of individuals and by that for allowing drawing decisions about the individual based on the assessment's results. Typical examples are questionnaires like the NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992), the Personality Research Form (PRF; Stumpf, Angleitner, Wieck, Jackson, & Beloch-Till, 1985), or the widely used but often criticized Minnesota Multiphasic Inventory (second and revised version MMPI-2 Butcher et al., 2000) or aptitude tests such as the Berlin Intelligence Structure Test (BIS; Beckmann, Guthke, Jäger, Süß, & Beauducel, 1999) or the Wechsler Adult Intelligence Scale (WAIS-IV; Petermann, 2012). The prior goal of their scale development was to maximize their reliability and validity, thereby increasing measurement precision and decreasing the risk of false decisions (Krueger, 2012). Survey instruments, by contrast, do not face this particular need. They are designed to measure differences within and between (sub-)populations; they are not intended to provide interpretable data on an individual level. However, in order to be able to make inferences about the population as a whole, large-scale surveys - as the name implies - must be based on large representative samples. This means that extensive efforts must be made to realize a sample that reflects the target population as accurately as possible. Therefore, random samples are usually drawn from sampling frames such as the population register, for example. Extensive efforts are made to achieve a high response rate (e.g., Dillman, 2007). Usually, target persons are contacted several times by letter, phone, or email; monetary or nonmonetary incentives conditional upon survey participation are offered - sometimes even unconditional incentives are given. Hence, a representative survey is designed in such a way as to maximize the response rate. Efforts are made to reduce or eliminate aspects of the questionnaire that adversely influence the likelihood of response. One of the most crucial aspects in this regard is the length of the questionnaire, which has been found to be negatively correlated with response rates (Edwards, Roberts, Sandercock, & Frost, 2004). Furthermore, assessment time in large-scale social surveys is expensive and severely limited. As a consequence, in order to increase the response rate and to keep costs at a reasonable level, only short-scale inventories can be applied in large-scale surveys.

Besides the severe limitation of assessment time, surveys usually aim to assess multiple themes and constructs. Despite the fact that psychological constructs have been shown to be powerful predictors of several socioeconomic outcomes, psychological scales are only one, often very minor, part of the survey questionnaire. Given these limitations, it is clear that a typical psychological questionnaire - for example, the NEO-PI-R (Costa & McCrae, 1992) with 240 items and an average completion time of over half an hour - is simply too lengthy and therefore not applicable in settings such as large-scale surveys.

In sum, even though there is a wide range of high quality questionnaires and tests for psychological constructs, most of them are too long to be used in large-scale surveys. Abbreviated or short scales for measuring the construct of interest are desirable alternatives.

Potential Limitations of Short Scales With Regard to Reliability and How to Handle Them

Faced with a distinct lack of appropriate and applicable instruments, survey researchers are often obliged to develop short-scale instruments themselves. This can give rise to two major problems. First, these ad hoc instruments have usually neither been published nor documented in any other form (Smith, McCarthy, & Anderson, 2000). Hence, survey researchers wishing

to assess a particular psychological construct usually start from scratch and develop a new instrument, which is therefore not comparable with those used in other surveys. For example, the Big Five dimensions of personality are assessed in the ISSP, the SOEP, the UK Household Panel, and the HILDA survey. However, all these surveys use different short-scale instruments for the assessment. The second, and more severe, problem is that, because short-scale measures are frequently constructed ad hoc by the survey researchers, the psychometric quality of the resulting instruments is often not comprehensively tested. Moreover, because these ad hoc measures are not usually developed by experts in psychological assessment, the quality of the instruments is quite often questionable.

Short-scale measures are frequently criticized for lacking psychometric quality (e.g., Krueger, Emons, & Sijtsma, 2013). This criticism is well founded, as pointed out above. Many of the widely used short-scale measures were developed from scratch and never thoroughly validated. Undoubtedly, like all other psychological measures, short scales have to meet specific psychometric quality standards in order to assure their interpretability (Burisch, 1984; Kersting, 2006; Nunnally, 1978).

Scale construction is still most commonly based on the true score theory, which is also known as the classical test theory (Lord & Novick, 1968). According to this theory, an observed test or questionnaire score comprises two parts: the measurement error and the true score. The true score refers to the actual individual score on the latent dimension of the construct of interest (Raykov & Marcoulides, 2011). As the classical test theory is fundamentally a theory of measurement error (Kaplan & Saccuzzo, 2009; Traub, 1997), one of the most crucial aspects in evaluating the quality of a scale is its reliability. Following classical test theory, the amount of random measurement error in the data can be minimized if responses across multiple indicators are averaged (Credé, Harms, Niehorster, & Gaye-Valentine, 2012). In turn, the reliability of the scale increases with the number of responses.

Most commonly, a scale's reliability is estimated using Cronbach's α (Cohen & Swerdlik, 2005; Kaplan & Saccuzzo, 2009; Niemi, Carmines, & McIver, 1986). Often regarded as an index of internal consistency (Boyle, 1991), Cronbach's α represents the average inter-item covariance of a set of items (Raykov & Marcoulides, 2011). Higher covariance between the items of a scale results in a higher α . To be interpreted correctly, Cronbach's α requires essential s-equivalent items, unidimensionality, and uncorrelated errors (Raykov & Marcoulides, 2011; Schermelleh-Engel & Werner, 2012; Sijtsma, 2009). Essential s-equivalence is given when the items of a scale measure the same true score of a construct while the error variances differ. In addition, "the essentially s-equivalent model allows each item true score to differ by an additive constant unique to each pair of variable" (Graham, 2006, p. 935). It has to be noted that these assumptions are rather too strong and restrictive for most psychological scales. In addition, Cronbach's α cannot be regarded as an index of unidimensionality but is still often interpreted as such (Raykov & Marcoulides, 2011; Sijtsma, 2009).

Compared to longer scales, lower levels of Cronbach's α are frequently reported for scales with few items (Schweizer, 2011). The problem is particularly relevant for short-scale construction. First, the aim of short-scale construction is to provide an economic measure with less redundant items while retaining the breadth of the construct of interest at the same time. When the construct of interest is a relatively broad one that encompasses numerous facets (e.g., Agreeableness, one of the Big Five personality dimensions), the item set of a short scale is usually selected in such a way that it reflects the breadth of the construct with a minimal number of indicators. As a consequence, the resulting scales contain (a) very few and (b) comparatively heterogeneous items measuring the construct in question. Therefore, the inter-item correlations of the short-scale items are small, thereby unsurprisingly yielding lower internal consistency compared to longer scales while the scale's validity in terms of content coverage is rather maintained. This problem has often been referred to as the attenuation paradox in measurement theory (Loevinger, 1954). The attenuation paradox describes the fact that under certain circumstances, enhancing the reliability of a scale may even go along with a decrease in validity. Second, coefficient α is known to be directly dependent on the number of (homogeneous) items in a scale. Thus, in the framework of classical test theory, "increasing the number of items [is] considered the proper means for increasing the reliability" (Schweizer, 2011, p. 71). (In the case of parallel indicators, the gain in reliability can be calculated by using the Spearman-Brown Prophecy Formula; Cohen & Swerdlik, 2005; Kaplan & Saccuzzo, 2009; Raykov & Marcoulides, 2011.) However, given the above-mentioned limitations of Cronbach's α , the extent to which the reliability of short scales is properly reflected by indicators of internal consistency must be questioned (Boyle, 1991; Raykov & Marcoulides, 2011; Sijtsma, 2009). Even for longer scales, coefficient α may yield unfavorable outcomes. Under certain circumstances, scales with a larger set of items are not necessarily more reliable than short scales. For example, when more heterogeneous items are added to a long scale, α decreases despite the higher number of items (Niemi et al., 1986). Schweizer (2011), on the other hand, emphasizes that short scales can show satisfactory levels of Cronbach's α even though they comprise a small number of items. In addition, the psychometric requirements with regard to reliability are different for survey instruments. According to Nunnally and Bernstein (1994), lower levels of Cronbach's α are acceptable for the investigation of group differences, whereas higher levels of the coefficient are required when it comes to drawing inferences about individual differences. As pointed out above, psychological short scales in social surveys are used for investigations on the group level rather than for individual assessment.

In sum, alternative estimates of reliability are required (Cronbach & Shavelson, 2004; Zinbarg, Revelle, Yovel, & Li, 2005). One possible option is to calculate McDonald's omega (McDonald, 1999; Raykov & Marcoulides, 2011), which describes the "proportion of variance in the scale scores accounted for by a general factor" (Zinbarg et al., 2005, p. 124). The main advantage of using McDonald's omega rather than Cronbach's α when evaluating the quality of short-scale measures is that reliability estimated using coefficient omega does not increase or decrease with the number of items in the scale. Further-

more, coefficient omega has fewer requirements than α (e.g., s-equivalence of indicators is not necessary; Zinbarg et al., 2005). In addition, it is usually estimated in a framework of confirmatory factor analysis, which provides an excellent opportunity to test the assumptions of Cronbach's α (e.g., unidimensionality, essential s-equivalence) as well as to estimate McDonald's omega.

Another strategy for estimating the reliability of a short scale is test-retest reliability (r_{tt}). Gosling, Rentfrow, and Swan (2003, p. 507) note that this strategy is particularly valuable for estimating the reliability of single-item scales as internal-consistency indices can in these cases not be computed (see also Heise, 1969, p. 93). Usually estimated by administering the scale twice and calculating the correlation between the two sets of scores, it can be applied provided both the true scores and the error variances are stable over time (Abell, Springer, & Kamata, 2009; Fishman & Galguera, 2003; Schermelleh-Engel & Werner, 2012).

As the test-retest reliability coefficient is calculated on the basis of scale scores, the number of items is not taken into account. For example, Gosling et al. (2003) demonstrated that short scales of Big Five measures can indeed achieve satisfactory test-retest reliability. Therefore, the reliability of short scales need not be automatically lower provided other, more appropriate, estimators of reliability are employed.

Potential Limitations of Short Scales With Regard to Validity and How to Handle Them

A scale's validity is at least as important as its reliability (Loevinger, 1954). Validity refers to "whether an instrument is indeed measuring what it purports to evaluate" (Raykov & Marcoulides, 2011, p. 183). When compared to scales with a larger set of items, short scales are often considered to be inferior in terms of validity (e.g., Credé et al., 2012; Niemi et al., 1986; Smith et al., 2000). There are several reasons for this. First, criterion validity is usually determined by measuring the correlation between the scale score (e.g., for self-efficacy) and the score on a criterion variable (e.g., academic success). According to classical test theory, the magnitude of the correlation between these two scores cannot exceed the product of their reliability indices (Kaplan & Saccuzzo, 2009; Raykov & Marcoulides, 2011). As outlined above, short scales have often been found to show lower reliability levels when indexed by Cronbach's α . Hence, the low reliability of a short scale may result in low criterion validity. Second, another expected deficiency of brief or abbreviated scales is related to their content validity (Credé et al., 2012; Smith et al., 2000). A small set of indicators cannot capture all facets of a broad personality trait. Some aspects of the construct of interest may be underrepresented in the scale. This is especially the case when items are selected on the basis of their item-total correlation, thereby yielding a set of highly homogeneous items (Kruyen et al., 2013). Consequentially, the statistics-driven selection strategy can also result in a set of highly redundant items (Boyle, 1991). Hence, the reduced coverage of the target domain severely limits the content validity of the short scale (Smith et al., 2000). This criticism applies in particular to single-item scales that tap only a single aspect of a construct (Credé et al., 2012).

As Schweizer (2011) points out, a satisfactory level of validity of a brief multi-item instrument can be also doubtful if a construct possesses a more complex factor structure (e.g., a hierarchical structure with higher-order factors like the Big Five). One way of dealing with this problem in short-scale construction is to reduce the number of constructs assessed by focusing only on the higher-order factors. For example, in a refined value circle model, Schwartz et al. (2012) describe 19 motivationally distinct values. The underlying motivations of the values can be summarized insofar as the 19 values load on four higher-order factors - namely, Conservation, Openness to Change, Self-Enhancement, and Self-Transcendence (Schwartz & Boehnke, 2004). In order to develop an instrument suitable for use in large-scale social surveys, Beierlein et al. (2014) selected items for the four higher-order factors instead of measuring the first-order factors (i.e., the 19 basic human values). As a consequence, the resulting short scale is limited to the assessment of the four higher-order dimensions. Hence, it does not allow conclusions to be drawn with regard to individual differences on the level of basic human values. Another strategy for dealing with the problem of content validity and to avoid the problem of reduced coverage is to deliberately focus on satisfactorily capturing the breadth of the construct in question when developing the short scale. By doing so, one consciously accepts a low level of internal consistency. For example, Clark and Watson (1995) recommend scale developers to follow a theory-driven approach of scale construction. From their point of view, striving for unidimensionality is superior to pursuing high levels of inter-item correlations. Also Kline (1986, cited by Boyle, 1991, p. 291) emphasizes that high levels of validity can be reached if not all items of a scale are highly interrelated with each other, but are positively associated with the criterion. Thus, short-scale developers should give special attention to validity concerns rather than focusing exclusively on reliability, in particular, on the level of internal consistency. This approach was taken by Rammstedt and John (2007), for example, when they were developing the BFI-10. Two items per Big Five dimension were selected to cover a maximal bandwidth

Table 1. Selected short scales assessing psychological constructs

Scale name	Reference	Number of Items per (Sub-)Scale	Reliability
Attractiveness Rating (AR1)	Lutz, Kemper, Beierlein, Margraf-Stiksrud, & Rammstedt (2013)	1	$r_{tt} = .46-.85$ (1-week interval)
12-Item Short-Form Health Survey (SF-12) - Mental Component Summary Scale	Ware, Kosinski, & Keller (1996)	1-2	$r_{tt} = .76$ (2-weeks interval)
Brief Measure of Sensation Seeking Scale (BSSS)	Hoyle, Stephenson, Palmgreen, Lorch, & Donohew (2002)	2	$\alpha = .74-.79$
Dirty Dozen (Narcissism, Machiavellianism, Psychopathy)	Jonason & Webster (2010)	4	$r_{tt} > .76$ (3-weeks interval)
Effort-Reward Imbalance - Short Scale (ERI-S)	Siegrist, Wege, Piehlhofer, & Wahrendorf (2009)	3-7	$\alpha > .70$
4-Item-Scale for the Assessment of Internal and External Control Beliefs (IE-4)	Kovaleva, Beierlein, Kemper, & Rammstedt (2012a)	2	$co = .53$ to $.71$; $f_{tt} = .56-.64$ (6-weeks interval)
General Self-Efficacy Short Scale (ASKU)	Beierlein, Kemper, Kovaleva, & Rammstedt (2013)	3	$co = .81-.86$; $r_{tt} = .50$ (6-weeks interval)
Justice Sensitivity	Baumert et al. (2013)	2	$co = .78-.92$; $f_{tt} = .44-.56$ (6-weeks interval)
Political Efficacy Short Scale (PEKS)	Beierlein et al. (2012a)	2	$co = .69-.92$; $r_{tt} = .44-.61$ (6-weeks average interval)
Scale Impulsive Behavior (1-8)	Kovaleva, Beierlein, Kemper, & Rammstedt (2012b)	2	$co = .65-.92$; $r_{tt} = .46-.57$ (6-weeks interval)
Scale Optimism-Pessimism-2 (SOP2)	Kemper, Beierlein, et al. (2013)	2	$co = .74-$, $.83$
Satisfaction With Life Scale (SWLS)	Diener, Emmons, Larsen, & Griffin (1985)	5	$r_{tt} = .82$ (2-months interval); $\alpha = .87$
Short Form of the Self-Compassion Scale (SCS-SF)	Raes, Pommier, Neff, & Van Gucht (2011)	2	$\alpha = .86$
Short Form of the State Scale of the Spielberger State-Trait Anxiety Inventory (STAI) "	Marteau & Bekker (1992)	6	$\alpha = .82$
Short Scale for the Assessment of crystallized intelligence (BEFKI GC-K)	Schipolowski, Wilhelm, Schroeders, Kovaleva, Kemper, & Rammstedt (2013)	12	$co = .82$; $\alpha = .81$
Short Scale Interpersonal Trust (KUSIV3)	Beierlein, Kemper, Kovaleva, & Rammstedt (2012b)	3	$co = .85$; $r_{tt} = .57$ (6-weeks interval)
Short version of the Big Five Inventory (BFI-10)	Rammstedt & John (2007)	2	$r_{tt} = .49-.84$ (6-8-weeks interval)
Single Item Self-Esteem Scale (SISE)	Robins, Hendin, & Trzesniewski (2001)	1	$r_{tt} = .61$ (across 6 assessments in 4 years)
Social Desirability Short Scale (KSE-G)	Kemper, Beierlein, Bensch, Kovaleva, & Rammstedt (2012)	3	$co = .78-.88$
Ten-Item Personality Inventory (TIPI)	Gosling et al. (2003)	2	$r_{tt} = .62-.71$ (6-weeks interval)

Notes. f_{tt} = Retest coefficient; co = McDonald’s Omega coefficient; α = Cronbach’s Alpha coefficient.

of the dimension. As a result, the scales are comparatively heterogeneous.

Contrary to the assumed negative effects of the brevity of a scale on its validity, several researchers have provided evidence that the validity of short scales is comparable to that of longer scales. For example, Robins, Hendin, and Trzesniewski (2001) demonstrated that their single-item measure of self-esteem achieves acceptable levels of convergent and criterion validity. Similar results are also shown for a brief measure of anxiety (Kemper, Lutz, & Neuser, 2011). Thalmayer, Saucier, and Eigenhuis (2011) compared the validity of several versions of the Big Five questionnaire of varying scale length. With regard to predictive validity, the more parsimonious scales performed equally well compared to scales with a larger number of items. Rammstedt and John (2007; see also Rammstedt, Kemper, Klein, Beierlein, & Kovaleva, 2013) report satisfactory levels of reliability and factorial, discriminant, congruent, and criterion validity for their ten-item Big Five measure.

The use of short scales may also have several other advantages with respect to validity. As Wanous, Reichers, and Hudy (1997) argue, the brevity of a scale may enhance face validity in the eyes of the respondent. As a short scale comprises only a small number of items that best represent the construct of interest, the respondent does not get the impression that the items in the scale are repeated. Moreover, a large number of items can have a negative impact on respondents' cognitive and motivational processes when answering a questionnaire (cf. Tourangeau, Rips, & Rasinski, 2000). For example, time-consuming questionnaires may cause boredom and fatigue (Burisch, 1984; Robins et al., 2001) and, by so doing, may seriously impair the respondent's motivation to fully complete the questionnaire. As a consequence, the amount of missing data increases. Furthermore, respondents may be inclined merely to fulfill the basic demands of survey participation. In order to minimize their efforts, they may choose the first available answer category rather than providing an optimal and accurate answer. In survey research, this phenomenon is commonly described as "satisficing" (Krosnick, 1999). Consequently, keeping questionnaires short increases the likelihood of obtaining optimally generated answers from the respondent (Burisch, 1984; Krosnick, 1999).

To meet the need for psychological short scales, and to increase the data quality and comparability of future surveys, initial attempts have been made to develop and provide well constructed and thoroughly tested short-scale measures for numerous psychological constructs (for collections of short scales, see, e.g., Kemper, Brahler, & Zenger, 2013; Rammstedt, Kemper, & Schupp, 2013).¹ These measures include short scales for assessing the Big Five personality dimensions (e.g., BFI-10; Rammstedt & John, 2007), optimism-pessimism (Kemper, Beierlein, Kovaleva, & Rammstedt, 2013), justice sensitivity (Baumert et al., 2013), and intelligence (e.g., Schipolowski et al., 2013). In Table 1, we present a selection of short scales of which several have been already used within the context of large-scale social surveys.

Conclusion

As shown above, there is an increasing interest in, and demand for, the assessment of key psychological constructs in large-scale surveys. This need can be met only by suitable short-scale measures. As few such measures exist at present, survey researchers are often obliged to develop short-scale measures by themselves - usually without the necessary expertise in this field. We have demonstrated that such short-scale measures do not automatically possess low psychometric properties. Rather, by using appropriate indicators, reliability and validity can be sufficient for the intended research objective, namely, comparisons within and across (sub-)populations.

References

- Abell, N., Springer, D. W., & Kamata, A. (2009). *Developing and validating rapid assessment instruments*. New York, NY: Oxford University Press.
- Allison, P. J., Guichard, C., Fung, K., & Gilain, L. (2003). Dispositional optimism predicts survival status 1 year after diagnosis in head and neck cancer patients. *Journal of Clinical Oncology*, *21*, 543-548.
- Arthur, W. Jr., & Graziano, W. G. (1996). The five-factor model, conscientiousness, and driving accident involvement. *Journal of Personality*, *64*, 593-618.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Baumert, A., Beierlein, C., Schmitt, M., Kemper, C. J., Kovaleva, A., Liebig, S., & Rammstedt, B. (2013). Measuring four facets of Justice Sensitivity with two items each. *Journal of Personality Assessment*, *1996*, 380-390. doi: 10.1080/00223891.2013.836526
- Beckmann, J. F., Guthke, J., Jäger, A. O., Süß, H.-M., & Beauducel, A. (1999). Berliner Intelligenzstruktur-Test (BIS), Form 4 [Berlin Intelligence Structure Test, Form 4]. *Diagnostica*, *45*, 55-61.
- Beierlein, C., Davidov, E., Behr, D., Cieciuch, J., László, Z., ... Schwartz, S. (2014). *Construction and validation of the survey compatible Schwartz Values Short Scale 4 using international samples*. Manuscript in preparation.
- Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2012a). *Ein Messinstrument zur Erfassung politischer Kompetenz- und Einflussserwartungen: Political Efficacy Kurzskala (PEKS)* [A measurement instrument for assessing political competence and control expectations: Political Efficacy Short Scale (PEKS)]. GESIS Working Papers 2012/118. Köln, Germany: GESIS.
- Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2012b). *Kurzskala zur Messung des zwischenmenschlichen Vertrauens: Die Kurzskala Interpersonales Vertrauen (KU-SIV3)* [Short scale for assessing interpersonal trust: The short scale interpersonal trust (KUSIV3)]. GESIS Working Papers 2012/122. Köln, Germany: GESIS.

- Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2013). Kurzsкала zur Erfassung allgemeiner Selbstwirksamkeitserwartungen (ASKU) [Short scale for assessing general self-efficacy expectations (ASKU)]. *Methoden, Daten, Analysen*, 7, 251-278.
- Borghans, L., Duckworth, A. L., Heckman, J. J., & Weel, B. T. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43, 972-1059.
- Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, 12, 291-294.
- Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*, 39, 214-227.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., Kaemmer, B., & Engel, R. (2000). *Minnesota Multiphasic Personality Inventory (MMPI-2)*. Göttingen, Germany: Hogrefe.
- Caprara, G. V., Vecchione, M., Capanna, C., & Mebane, M. (2009). Perceived political self-efficacy: Theory, assessment, and applications. *European Journal of Social Psychology*, 39, 1002-1020.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309-319.
- Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurement*. Boston, MA: McGraw-Hill Publishing.
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five Factor Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Crede, 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, 874-888.
- Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391-418.
- Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49, 71-75.
- Dillman, D. A. (2007). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York, NY: Wiley.
- Edwards, P., Roberts, I., Sandercock, P., & Frost, C. (2004). Follow-up by mail in clinical trials. Does questionnaire length matter? *Controlled Clinical Trials*, 25, 31-52.
- Fishman, J. A., & Galguera, T. (2003). *Introduction to test construction in the social and behavioral sciences: A practical guide*. Ranham, MD: Rowman & Littlefield.
- German Data Forum. (2010). *Building on progress: Expanding the research infrastructure for the social, economic, and behavioral sciences*. Opladen, Germany: Budrich UniPress.
- Goldberg, L. R. (2005). *Why personality measures should be included in epidemiological surveys: A brief commentary and a reading list*. Eugene, OR: Oregon Research Institute.
- 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, 504-528.
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24, 79-132.
- Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, 13, 1-4.
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*, 66, 930-944.
- Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, 34, 93-101.
- Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences*, 32, 401-414.
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, 22, 420.
- Kaplan, R., & Saccuzzo, D. (2009). *Psychological testing: Principles, applications, and issues*. Belmont, CA: Wadsworth, Cengage Learning.
- Kemper, C. J., Beierlein, C., Bensch, D., Kovaleva, A., & Rammstedt, B. (2012). *Eine Kurzsкала zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens: Die Kurzsкала Soziale Erwünschtheit-Gamma (KSE-G)* [A short scale for assessing the gamma-factor of social desirable response behavior: The short scale Social Desirability-Gamma (KSE-G)]. GESIS Working Papers 2012I25. Köln, Germany: GESIS.
- Kemper, C. J., Beierlein, C., Kovaleva, A., & Rammstedt, B. (2012). *Eine Kurzsкала zur Messung von Optimismus-Pessimismus: Die Skala Optimismus-Pessimismus-2 (SOP2)* [A Short Scale for Assessing Optimism-Pessimism: The Optimism-Pessimism-2 Scale (SOP2)]. GESIS Working Papers 2012I15. Köln, Germany: GESIS.
- Kemper, C. J., Beierlein, C., Kovaleva, A., & Rammstedt, B. (2013). Entwicklung und Validierung einer ultrakurzen Operationalisierung des Konstrukts Optimismus-Pessimismus-Die Skala Optimismus-Pessimismus-2 (SOP2) [Development and validation of an ultra-short assessment of optimism-pessimism-The Optimism-Pessimism-2 Scale (SOP2)]. *Diagnostica*, 59, 119-129.
- Kemper, C. J., Brähler, E., & Zenger, M. (Eds.). (2013). *Psychologische und sozialwissenschaftliche Kurzsкаlen: Standardisierte Erhebungsinstrumente für Wissenschaft und Praxis* [Psychological and social science short scales: Standardized instruments for research and practice]. Berlin, Germany: Medizinisch-Wissenschaftliche Verlagsgesellschaft.
- Kemper, C. J., Lutz, J., & Neuser, J. (2011). Konstruktion und Validierung einer Kurzform der Skala Angst vor negativer Bewertung (SANB-5) [Construction and Validation of a scale for assessing fear of negative appraisal]. *Klinische Diagnostik und Evaluation*, 4, 343-360.
- Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn [The evaluation of test quality: Summary and new beginning]. *Psychologische Rundschau*, 57, 243-253.
- Kovaleva, A., Beierlein, C., Kemper, C. J., & Rammstedt, B. (2012a). *Eine Kurzsкала zur Messung von Kontrollüberzeugung: Die Skala Internale-Externale-Kontrollüberzeugung-4 (IE-4)* [A short scale for assessing control beliefs: The scale Internal-External Control Beliefs-4 (IE-4)]. GESIS Working Papers 2012I19. Köln, Germany: GESIS.
- Kovaleva, A., Beierlein, C., Kemper, C. J., & Rammstedt, B. (2012b). *Eine Kurzsкала zur Messung von Impulsivität nach dem UPPS-Ansatz: Die Skala Impulsives-Verhalten-5 (I-5)* [A short scale for assessing impulsivity following the UPPS approach]. GESIS Working Papers 2012I20. Köln, Germany: GESIS.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50, 537-567.
- Kruyen, P. M. (2012). *Using Short Tests and Questionnaires for Making Decisions about Individuals: When is Short Too Short?* (Doctoral

- dissertation, Tilburg University). Retrieved from: <http://arno.uvt.nl/show.cgi?fid=128226>
- Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, 13, 223-248.
- Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin*, 51, 493-504.
- Lohmann, H., Spieß, C. K., Groh-Samberg, O., & Schupp, J. (2009). Analysepotenziale des sozio-oekonomischen Panels (SOEP) für die empirische Bildungsforschung [Analytic potential of the German Socio-economic Panel for educational research]. *Zeitschrift für Erziehungswissenschaft*, 12, 252-280.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theory of mental test scores*. Reading, MA: Addison-Wesley.
- Lutz, J., Kemper, C. J., Beierlein, C., Margraf-Stiksrud, J., & Rammstedt, B. (2013). Konstruktion und Validierung einer Skala zur relativen Messung von physischer Attraktivität mit einem Item: Das Attraktivitätsrating 1 (AR1) [Construction and validation of a scale for the relative assessment of physical attractiveness using a single item: The attractiveness rating 1 (AR1)]. *Methoden, Daten, Analysen*, 7, 209-232.
- Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, 31, 301-306.
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Mahwah, NJ: Erlbaum.
- Mondak, J. J. (2010). *Personality and the Foundations of Political Behavior*. New York, NY: Cambridge University Press.
- Niemi, R. G., Carmines, E. G., & McIver, J. P. (1986). The impact of scale length on reliability and validity. *Quality and Quantity*, 20, 371-376.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Petermann F. (Ed.). (2012). *Wechsler Adult Intelligence Scale- Fourth Edition (WAIS-IV)* Deutsche Version [German Version]. Frankfurt, Germany: Pearson Assessment.
- Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the self-compassion scale. *Clinical Psychology & Psychotherapy*, 18, 250-255.
- Rammstedt, B. (2010). Subjective indicators. In German Data Forum. (Eds.), *Building on progress. Expanding the research infrastructure for the social, economic, and behavioral sciences* (pp. 813-824). Opladen: Budrich UniPress.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41, 203-212.
- Rammstedt, B., Kemper, C. J., & Schupp, J. (Eds.). (2013). Standardisierte Kurzskalen zur Erfassung psychologischer Merkmale in Umfragen. [Standardized short-scale measures for assessing psychological constructs in surveys]. *Methoden, Daten, Analysen*, 7, 145-152.
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: 10 Item Big Five Inventory (BFI-10) [A short scale for assessing the Big Five dimensions of personality (BFI-10)]. *Methoden, Daten, Analyse*, 7, 235-251.
- Rammstedt, B., & Spinath, F. M. (2013). Öffentliche Datensätze und ihr Mehrwert für die psychologische Forschung [Publicly available data sets and their use for psychological research]. *Psychologische Rundschau*, 64, 101-102.
- Rasmussen, H. N., Scheier, M. F., & Greenhouse, J. B. (2009). Optimism and physical health: A meta-analytic review. *Annals of Behavioral Medicine*, 32, 239-256.
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Taylor & Francis.
- 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*, 22, 151-161.
- Schermerle-Engel, K., & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung [Methods of estimating reliability]. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* [Test and questionnaire construction] (2nd ed., pp. 119-142). Heidelberg, Germany: Springer.
- Schipolowski, S., Wilhelm, O., Schroeders, U., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2013). BEFKIGC-K: Eine Kurzskala zur Messung kristalliner Intelligenz [BEFKI GC-K: A short scale for the measurement of crystallized intelligence]. *Methoden, Daten, Analysen*, 7, 155-183.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schoen, H., & Steinbrecher, M. (2013). Beyond total effects: Exploring the interplay of personality and attitudes in affecting turnout in the 2009 German federal election. *Political Psychology*, 34, 533-552.
- Schupp, J., Spieß, C. K., & Wagner, G. G. (2008). Die verhaltenswissenschaftliche Weiterentwicklung des Erhebungsprogramms des SOEP [The behavioral research enhancements of the SOEP survey programme]. *Vierteljahreshefte zur Wirtschaftsforschung*, 22, 63-76.
- Schwartz, S. H., & Boehnke, K. (2004). Evaluating the structure of human values with confirmatory factor analysis. *Journal of Research in Personality*, 38, 230-255.
- Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., ... Konty, M. (2012). Refining the theory of basic individual values: New concepts and measurements. *Journal of Personality and Social Psychology*, 103, 663-688.
- Schweizer, K. (2011). Some thoughts concerning the recent shift from measures with many items to measures with few items. *European Journal of Psychological Assessment*, 22, 71-72. doi: 10.1027/1015-5759/a000056
- Siegrist, J., Wege, N., Pühlhofer, F., & Wahrendorf, M. (2009). A short generic measure of work stress in the era of globalization: Effort-reward imbalance. *International Archives of Occupational and Environmental Health*, 82, 1005-1013.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 24, 107-120.
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short form development. *Psychological Assessment*, 12, 102-111.
- Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35, 401-426.
- Stumpf, H., Angleitner, A., Wicke, T., Jackson, D. N., & Beloch-Till, H. (1985). *Deutsche Personality Research Form (PRF). Handanweisung (Manual)*. Göttingen, Germany: Hogrefe.
- Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief- to medium-length Big Five and Big Six questionnaires. *Psychological Assessment*, 23, 995-1009.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
- Traub, R. (1997). Classical Test Theory in historical perspective. *Educational Measurement: Issues and Practice*, 16, 8-14.
- Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology*, 82, 247-252.
- Ware, J. E. Jr., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability

and validity. *Medical Care*, **34**, 220-233.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's Alpha, Revelle's Beta, McDonald's Omega: Their relations with each and two alternative conceptualizations of reliability. *Psychometrika*, **70**, 123-133.

Date of acceptance: June 12, 2014 **Published online:** November 21, 2014

Beatrice Rammstedt

GESIS - Leibniz Institute for the Social Sciences PO Box 12 21 55 68072 Mannheim Germany

E-mail beatrice.rammstedt@gesis.org