Further Considerations in Using Items With Diverse Content to Measure Acquiescence

Educational and Psychological
Measurement

2016, Vol. 76(1) 164–174

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164415586831

epm.sagepub.com

Chester Chun Seng Kam¹

Abstract

To measure the response style of acquiescence, researchers recommend the use of at least 15 items with heterogeneous content. Such an approach is consistent with its theoretical definition and is a substantial improvement over traditional methods. Nevertheless, measurement of acquiescence can be enhanced by two additional considerations: first, to ensure that it is not confounded with item valence, equal numbers of positively and negatively phrased items should be used; second, the validity of acquiescence scores can be substantially improved when derived from more than 15 items: about twice that number appears to be the minimum. These issues are illustrated empirically and recommendations are made for research practice.

Keywords

acquiescence, measurement, response style

Survey response styles can distort construct means and variance, inflate or deflate correlations, and alter factor analytic results. In short, they can lead to substantial bias in construct measurement and in statistical findings (Kam & Meyer, 2015a; Rammstedt & Farmer, 2013). For this reason, researchers often try to identify and control for them (Bentler, Jackson, & Messick, 1971; Greenleaf, 1992; Van Vaerenbergh & Thomas, 2013).

Unfortunately, methodologists do not agree on how to measure response styles, including acquiescence (De Beuckelaer, Weijters, & Rutten, 2010). As a result,

¹University of Macau, Taipa, Macau, China

Corresponding Author:

Chester Chun Seng Kam, Faculty of Education, University of Macau, E33 Avenida da Universidade, Taipa, Macau, China.

Email: ChesterKam@umac.mo

researchers may be using suboptimal—if not inaccurate—methods to measure it. The purpose of this article is to suggest better ways of measuring acquiescence bias. The work has implications both for methodologists who study response bias and applied researchers who want to measure and control for it.

Measurement of Acquiescence Response Style

Acquiescence may be defined as participants' tendency to agree or disagree with survey items regardless of the actual content (Bentler et al., 1971). It is usually measured by one of three methods (see Van Vaerenbergh & Thomas, 2013, for a review). In the first, item scores across the entire survey are summed. Although it is assumed that aggregate scores from a diverse set of scales would not measure any meaningful content (Schimmack, Oishi, & Diener, 2005), that assumption has been shown to be questionable (De Beuckelaer et al., 2010; Weijters, Baumgartner, & Schillewaert, 2013). Therefore, this method may confound acquiescence with substantive construct content. In the second method of measurement, scores on items with opposite content (e.g., "I am worthy"; "I am worthless") are summed (e.g., Rammstedt & Farmer, 2013; Winkler, Kanouse, & Ware, 1982). This method, however, confounds acquiescence with genuine attitudinal ambivalence, which is especially prevalent in collectivistic cultures (Peng & Nisbett, 1999). In the third method, items with heterogeneous content are summed (e.g., Baumgartner & Steenkamp, 2001). De Beuckelaer et al. (2010; see also Weijters et al., 2013) argued persuasively that this method is superior to the other two. Even so, two additional issues have not been fully addressed, potentially leading to suboptimal, if not inaccurate, measurement of acquiescence.

First Issue: Confounding With Item Valence

Based on previous discussions (Beauvois & Dubois, 2000; Peabody, 1967; Saucier, 1997), Kam and Meyer (2012; Kam & Meyer 2015b) argued that a survey item is seldom neutral. Rather, it has a certain valence. An item's objective content is its *descriptive* aspect; the item's valence (or favorability) is its *evaluative* aspect. Consider even a relatively "objective" self-report item, such as how often one is late for work (with Likert-type scale anchors from *Never* to *Always*). Whereas the descriptive aspect is the factual frequency of lateness, the evaluative aspect is the person's subjective opinion of lateness (e.g., How bad is it to be late for work?). As such, it is vulnerable to social desirability response style (Kam & Meyer, 2015b). Descriptive and evaluative aspects can seldom, if ever, be separated. Therefore, although ideally acquiescence should be measured using items with neutral content, in practice this is not feasible because most survey items have either positive or negative valence (Bäckström, Björklund, & Larsson, 2009; Peabody, 1967).

Previously, when researchers sampled items with heterogeneous content to measure acquiescence, they did not take valence into account. If they included items that are predominantly positively or negatively valenced, the resulting acquiescence score

may reflect participants' response bias to positive or negative items. To our knowledge, the issue of item valence has not been addressed in the acquiescence measurement literature.

Second Issue: Weak Convergent Validity

De Beuckelaer et al. (2010) found very low correlation ($r = .14$) between two acquiescence indicators made up of 15 heterogeneous items in the same survey. The authors concluded, "The low correlation suggests that little ARS [acquiescence response style] variance spills over from one set of items to the other" (p. 771). The result suggests that measures of acquiescence can be unstable and show very weak convergent validity.

However, this conclusion is at odds with other research demonstrating substantial stability of acquiescence scores within the same survey and over time. Billiet and McClendon (2000) found that acquiescence factors from two distinct scales in the same survey were moderately correlated ($r = .44$), while Billiet and Davidov (2008) found that acquiescence factors were correlated ($r = .56$) over a 4-year period. Similarly, acquiescence factors were stable over a 1-year period (Weijters, Geuens, & Schillewaert, 2010b). Furthermore, although Weijters, Geuens, and Schillewaert (2010a) did not directly report the correlation of acquiescence indicators within the same survey, they did find sizable averaged factor loadings (.67) when acquiescence indicators loaded on the same latent factor.

How can one reconcile the weak correlation of acquiescence scores within the same survey (De Beuckelaer et al., 2010) with their substantial stability in others (Billiet & McClendon, 2000; Weijters et al., 2010a, 2010b)?

The weak convergent validity in De Beuckelaer et al. (2010) was possibly due to the small number of items. As mentioned, Weijters et al. (2010a) found a good relationship (average $r = .67$) between the latent factor of acquiescence and a set of its observed indicators (each made up of 22 or 23 items). However, the relationship dropped substantially when the number of composite items was reduced: The average correlation between the latent factor of acquiescence and its set of indicators shrank to .22 when an indicator was made up of only one item. This finding clearly suggests that the stability of an acquiescence score depends on its number of composite items. Therefore, the weak validity found by De Beuckelaer et al. (2010) may potentially be overcome with a larger number of items.

The Present Research

The present study investigated two hypotheses: (a) If an acquiescence scale has an imbalance of positively and negatively phrased items, acquiescence will be confounded with item valence; (b) the convergent validity of an acquiescence score will increase with increasing number of items used to derive it.

To test the first hypothesis, three averaged item scores were compared: one from positively phrased items only, one from negatively phrased items only, and one from

equal numbers of positively and negatively phrased items. Each score comprised 16 items. If the first two scores (i.e., positively or negatively phrased items only) are confounded with item valence, they should be negatively correlated. Furthermore, because the first two scores have items with identical valence, their interitem correlations are expected to be stronger than the third score. Finally, if the first two scores are confounded with valence, they should—relative to the third, balanced score—have weaker convergent validity with other measures of acquiescence.

To test the second hypothesis, four acquiescence scores were created—comprising 16, 32, 64, and 128 items—and correlated with two external indicators. The first was based on the summed scores of items with opposite content. Although, as noted earlier, this computational method may be confounded with attitudinal ambivalence, the sample used was dominated by Caucasian students, who tend to be low in ambivalent attitude (Peng & Nisbett, 1999). The second external indicator was based on the response style factor scores computed from random intercept item factor analysis (RIIFA; Maydeu-Olivares & Coffman, 2007). Maydeu-Olivares and Coffman (2007) have argued that their factor analytic model can reasonably capture acquiescence. Kam and Zhou (2015) showed that such an acquiescence score has moderate validity in real data. Finally, the large pool of items (128) allowed two versions of the 16-, 32-, and 64-item scores to be created. The intercorrelations between two versions of the same score were examined. High correlations would be further support for convergent validity (Billiet & McClendon, 2000; Weijters et al., 2010a, 2010b).

Method

Introductory psychology students ($N = 1,023$; 664 females, 351 males, 8 unidentified; $M_{\text{age}} = 18.43$, $SD_{\text{age}} = 1.89$) completed an online mass testing survey at a large Canadian university in exchange for course credit. The university population was predominantly Caucasian. Data from the survey have been used by us (Kam & Meyer, 2012; Kam & Zhou, 2015) and other researchers in a variety of studies not directly related to the current investigation.

Acquiescence Indicators From Diverse Content

Indices of acquiescence were calculated by sampling a large pool of items with diverse content, ranging from personality and social psychology to clinical psychology. Sample items include the following: “Some groups of people are just more worthy than others” from the Social Dominance Orientation Scale (Sidanius & Pratto, 2001) and “I frequently get so strongly absorbed in one thing that I lose sight of other things” from the Autism Spectrum Quotient (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). The original scaling of the items is from *Strongly Disagree* to *Strongly Agree*; the number of scale points varies. Each item was rescaled so that it ranged from 0 to 1.

Figure 1. Random Intercept Item Factor Analysis.

Note. y_1 to y_5 = regular-keyed items; y_6 to y_{10} = reverse-keyed items; η_1 = latent variable measuring substantive content; η_2 = random intercept measuring acquiescence response style. The effect of acquiescence (η_1) on each measurement item is assumed to be identical. The variance of η_1 is fixed as 1; the variance of η_2 is freely estimated.

Calculation of Acquiescence Scores. Acquiescence Indicators From Diverse Content16-positive1 (AIDC16-positive1) and AIDC16-positive2 were calculated from two separate sets of 16 items with positive (favorable) content (e.g., “I’m always optimistic about my future”). Similarly, AIDC16-negative1 and AIDC16-negative2 were made up of 16 items with negative (unfavorable) content (e.g., “I feel anxious when having to use the tables in the back of a math book”). AIDC16-balanced1 and AIDC16-balanced2 were made up of 16 items, half with positive content and half with negative content. To examine the effect of number of items on acquiescence indicators, the number of items in the two AIDC16-balanced indicators was doubled to create AIDC32-balanced1 and AIDC32-balanced2, and doubled again to create AIDC64-balanced1 and AIDC64-balanced2. Finally, the two 64-item indicators were combined to create a 128-item score (AIDC128-balanced).

External Indicators. Two sets of external acquiescence scores were calculated to validate the AIDC scores. The first was calculated by summing the raw scores of five personality measures (NEO domain; Goldberg et al., 2006) to derive five separate indices of acquiescence. In addition, the raw scores of all five personality factors were averaged to form an overall acquiescence indicator. The second set of external scores was estimated by RIIFA analysis (Maydeu-Olivares & Coffman, 2007; see

Figure 1). Kam and Zhou (2015) showed that RIIFA-extracted scores, particularly when averaged from multiple distinct measures, have moderate validity in estimating acquiescence (see also Billiet & McClendon, 2000). Individual RIIFA acquiescence scores were thus estimated for each of the personality scales (Goldberg et al., 2006) before being summed to derive a combined RIIFA acquiescence index. All RIIFA acquiescence scores could be properly estimated except the one from the openness measure, which is thus excluded from the analysis.

Results

Analyses were conducted with R 3.0.2 (R Core Development Team, 2013). Missing data were treated with multiple imputation procedure from the Amelia II package in R. RIIFA requires the use of confirmatory factor analysis, which was conducted with lavaan package (with robust maximum likelihood estimator) in R.

Discriminant Validity Evidence

The averaged interitem correlations in AIDC16 indices were estimated (see Table 1). As predicted, the acquiescence scores calculated with positive (AIDC16-positive1 and AIDC16-positive2) and negative items (AIDC16-negative1 and AIDC16-negative2) apparently had stronger interitem correlations ($r_s = .09, .11, .05$, and $.05$) than the balanced scores (AIDC16-balanced1 and AIDC16-balanced2; $r_s = .009, .04$). This result shows that the positive and negative indices were more likely to capture some homogeneous quality—such as item valence—than were the balanced indices. The intercorrelations among AIDC-16 indices were then examined. An obvious finding is that the indices from positive items correlated negatively with indices from negative items ($r_s = -.34$ to $-.43$), showing discriminant validity. This finding demonstrates that the two sets of acquiescence indices captured item valence in addition to acquiescence.

Convergent Validity Evidence

The correlation of AIDC indices with opposite-item acquiescence indices are shown in Table 2. It is evident that AIDC indices did not correlate very well with opposite-item acquiescence indices measured by individual scales. The correlation was higher, however, for opposite-item acquiescence scores that are summed across individual scales (right-hand column). Another noticeable finding is that the correlations with combined opposite-item acquiescence indices were higher for AIDC-balanced indices (i.e., AIDC16-balanced1 and -balanced2) than for AIDC-positive or -negative indices (i.e., AIDC16-positive1, -positive2, -negative1, and -negative2). This result is particularly apparent for the combined opposite-item acquiescence index (right-hand column). Finally, AIDC-balanced indices that contain a larger number of items showed stronger convergent validity. This result is also particularly apparent for the combined

Table 1. Correlations Among AIDC16-Positive, AIDC16-Negative, and AIDC16-Balanced Indices.

AIDC Indices	Cronbach's α	Averaged interitem r	1	2	3	4	5	6
1. AIDC16-positive1	.51	.09	—					
2. AIDC16-positive2	.59	.11	.70	—				
3. AIDC16-negative1	.42	.05	-.38	-.43	—			
4. AIDC16-negative2	.41	.05	-.34	-.39	.62	—		
5. AIDC16-balanced1	.12	.009	.23	.26	.25	.17	—	
6. AIDC16-balanced2	.32	.04	.24	.13	.22	.22	.23	—

Note. $N = 1,023$. AIDC = acquiescence indicators from diverse content. For correlations between AIDC16 indices, all $ps < .001$.

Table 2. Correlations of AIDC Indices With Opposite-Item Acquiescence Indices.

AIDC indices	Cronbach's α	Averaged interitem r	Opposite-item acquiescence indices					
			Openness Acquiescence	Conscientiousness Acquiescence	Extraversion Acquiescence	Agreeableness Acquiescence	Neuroticism Acquiescence	Combined Acquiescence
Nonbalanced								
AIDC16-positive1	.51	.09	.19	.15	.15	.26	.02	.25
AIDC16-positive2	.59	.11	.18	.14	.10	.31	-.05	.22
AIDC16-negative1	.42	.05	.20	.33	.20	.15	.28	.37
AIDC16-negative2	.41	.05	.16	.21	.27	.16	.33	.36
Balanced								
AIDC16-balanced1	.12	.009	.40	.36	.16	.23	.14	.41
AIDC16-balanced2	.32	.04	.16	.28	.35	.33	.38	.49
AIDC32-balanced1	.26	.01	.36	.39	.30	.35	.28	.54
AIDC32-balanced2	.25	.009	.29	.38	.33	.34	.31	.54
AIDC64-balanced1	.47	.02	.41	.48	.36	.42	.32	.64
AIDC64-balanced2	.44	.01	.36	.41	.39	.46	.32	.63
AIDC128-balanced	.59	.02	.42	.49	.41	.48	.35	.70

Note. $N = 1,023$. AIDC = acquiescence indicators from diverse content. For correlations between AIDC indices and opposite-item acquiescence indices that are at .10 or above, $ps < .001$.

Table 3. Correlations of AIDC Indices With RIIFA Acquiescence Indices.

AIDC indices	RIIFA acquiescence indices				
	Conscientiousness Acquiescence	Extraversion Acquiescence	Agreeableness Acquiescence	Neuroticism Acquiescence	Combined Acquiescence
Nonbalanced					
AIDC16-positive1	.23	.23	.37	.23	.38
AIDC16-positive2	.25	.21	.41	.15	.40
AIDC16-negative1	.23	.13	.04	.10	.19
AIDC16-negative2	.10	.21	.06	.19	.17
Balanced					
AIDC16-balanced1	.36	.17	.25	.16	.37
AIDC16-balanced2	.25	.34	.34	.30	.43
AIDC32-balanced1	.39	.33	.38	.29	.51
AIDC32-balanced2	.36	.34	.35	.29	.49
AIDC64-balanced1	.47	.38	.44	.34	.60
AIDC64-balanced2	.41	.43	.48	.36	.61
AIDC128-balanced	.48	.44	.50	.38	.66

Note. $N = 1,023$. AIDC = acquiescence indicators from diverse content; RIIFA = random intercept item factor analysis. For correlations between AIDC indices and RIIFA acquiescence indices that are at .10 or above, $ps < .001$.

index. Similar results were obtained when AIDC indices were correlated with RIIFA acquiescence scores (Table 3).

Finally, as each AIDC16-, AIDC32-, and AIDC64-balanced index had two versions, the correlations between versions were examined. The correlation between the two AIDC16 versions was low, $r = .23$, $p < .001$. However, it substantially increased with increased number of items: AIDC32, $r = .50$, $p < .001$; AIDC64, $r = .67$, $p < .001$. De Beuckelaer et al. (2010) attributed the low correlation between two versions of a 15-item index to the possible instability of acquiescence among participants. On the contrary, the current analysis shows considerable stability in acquiescence scores, provided a large number of items are used.

Discussion and Conclusion

De Beuckelaer et al. (2010) correctly advocated the use of heterogeneous item content to measure acquiescence response style. The purpose of the current research is to further elaborate the requirements for accurate measurement. Two issues have been inadequately addressed in the past.

The first is item valence. If the number of positively and negatively phrased items is not balanced, acquiescence will be confounded with item valence. This was shown by negative correlations between an acquiescence score measured exclusively by positively phrased items and another measured exclusively by negatively phrased items. Furthermore, the valence-confounded scores had generally lower correlations with another external indicator of acquiescence (opposite-item acquiescence indices) than did scores derived from balanced scales. To measure acquiescence, researchers commonly pick items at random from various scales. However, because most measurement instruments lack reverse-keyed items, randomly chosen items will probably not be balanced in valence. If a survey measures a positive construct such as engagement or extraversion, the resulting acquiescence scores may therefore be confounded with positive-item valence.

The second issue is the number of items required to attain desirable validity. Using only 16 items, the present investigation did not find the score to achieve an adequate level of validity with combined acquiescence scores (right-hand column in Tables 2 and 3), $r_s = .37$ to $.49$. Increasing the number of items to 32 raised the convergent validity to $.49$ to $.54$; 64 items raised validity to $.60$ to $.64$; and 128 items raised it to $.66$ to $.70$. If applied researchers are interested in obtaining a reasonable measure of acquiescence, it appears that using 32 items is the minimum required to achieve convergent validity of approximately $.50$. However, methodological researchers who wish to measure acquiescence more accurately may choose to use 64 or even 128 items to attain coefficients in the $.60$ to $.70$ range. Using a large number of items to measure acquiescence is not unheard of: Weijters et al. (2010b) used 112 in one sample. In any case, the investigation here clearly suggests that a large number of items are required to achieve good validity.

Author's Note

The data were collected when the author was a graduate student at the Department of Psychology, The University of Western Ontario.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was financially supported by a research grant from the University of Macau (SRG2013-00050-FED).

References

- Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality, 43*, 335-344.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*, 241-251.
- Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*, 143-156.
- Beauvois, J. L., & Dubois, N. (2000). Affordances in social judgment: Experimental proof of why it is a mistake to ignore how others behave towards a target and look solely at how the target behaves. *Swiss Journal of Psychology, 59*, 16-33.
- Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin, 76*, 186-204.
- Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research, 36*, 542-562.
- Billiet, J. B., & McClendon, M. K. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling, 7*, 608-628.
- De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity, 44*, 761-775.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.
- Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly, 56*, 328-351.
- Kam, C., & Meyer, J. P. (2012). Do optimism and pessimism have different relationships with personality dimensions? A re-examination. *Personality and Individual Differences, 52*, 123-127.

- Kam, C. C. S., & Meyer, J. P. (2015a). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*. Advance online publication. doi:10.1177/1094428115571894
- Kam, C. C. S., & Meyer, J. P. (2015b). Implications of item keying and item valence for the investigation of construct dimensionality. *Multivariate Behavioral Research*. Advance online publication. doi:10.1080/00273171.2015.1022640
- Kam, C. C. S., & Zhou, M. (2015). Does acquiescence affect individual items consistently? *Educational and Psychological Measurement*, 75, 764-784.
- Maydeu-Olivares, A., & Coffman, D. L. (2007). Random intercept item factor analysis. *Psychological Methods*, 11, 344-362.
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology Monograph*, 7, 1-18.
- Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, 54, 741-754.
- R Core Development Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, 25, 1137-1145.
- Saucier, G. (1997). Effects of variable selection on the factor structure of person descriptors. *Journal of Personality and Social Psychology*, 73, 1296-1312.
- Schimmack, U., Oishi, S., & Diener, E. (2005). Individualism: A valid and important dimension of cultural differences between nations. *Personality and Social Psychology Review*, 9, 17-31.
- Sidanius, J., & Pratto, F. (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge, England: Cambridge University Press.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25, 195-217.
- Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, 18, 320-334.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, 34, 105-121.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The stability of individual response styles. *Psychological Methods*, 15, 96-110.
- Winkler, J. D., Kanouse, D. E., & Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, 67, 555-561.