

# Dealing with Common Method Variance in International Marketing Research

Hans Baumgartner and Bert Weijters

Journal of International Marketing

2021, Vol. 29(3) 7–22

© American Marketing Association 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1069031X21995871

journals.sagepub.com/home/jim



## Abstract

Common method variance (CMV) is an important concern in international marketing research because presumed substantive relationships may actually be due to shared method variance. Because method effects may vary systematically across cultures and countries, accounting for method effects in international marketing research is particularly critical. A systematic review of *Journal of International Marketing* articles published during a five-year period (2015–2019,  $N = 93$ ) shows that (1) authors often report post hoc CMV tests but usually conclude that CMV is not an issue and (2) many post hoc tests are conducted using the Harman one-factor test and the marker variable technique, which have serious deficiencies for detecting and controlling CMV. Drawing on a classification and comparative evaluation of the most common statistical approaches for dealing with CMV, the authors recommend two approaches and propose a procedure for dealing with CMV in international marketing research. The procedure, which is based on multisample structural equation modeling, is illustrated with data from a cross-national pan-European survey ( $N = 11,970$ , 14 countries), which shows that even though method variance is present in the data, method effects do not seriously bias the substantive conclusions in this particular study.

## Keywords

common method variance, international survey research, method bias, method effects, multisample structural equation modeling

There is broad consensus among methodological researchers that random and systematic measurement error can seriously distort observed responses (Baumgartner and Weijters 2019; Podsakoff et al. 2003; Viswanathan 2005). While the use of multiple measures to assess a construct of interest can mitigate the adverse effects of random measurement error, systematic measurement error poses more difficult challenges. A common source of systematic measurement error is method variance. Method variance is present when an observed response reflects not only the construct the researcher intended to measure (and usually random measurement error) but also the method of measurement on which the observed responses were based. If method variance is not accounted for appropriately, it can be mistaken for substantive variance, and the conclusions derived from the research might be misleading. This situation is particularly serious when multiple measures of the same construct or different constructs share the same measurement method, which gives rise to common method variance (CMV).

Although CMV is a problem in any research in which non-substantive systematic influences affect observed responses, it may be expected to be of particular concern in international

marketing research. The reason is that researchers often want to either (1) establish the cross-cultural or cross-national invariance of construct scores or relationships between constructs or (2) demonstrate predicted differences between cultures and countries. If there are systematic differences in CMV across groups, cross-cultural and cross-national comparisons that do not take method effects into account may yield misleading conclusions about invariance or predicted differences in construct scores or relationships between constructs. For example, Wong, Rindfleisch, and Burroughs (2003) found cross-cultural differences in method effects related to reverse-worded items; Van Auken, Barry, and Bagozzi (2006) found cross-cultural variation in the way respondents used different scale formats when measuring cognitive age; and Tellis and Chandrasekaran (2010) found cross-cultural differences in the extent to which response biases (such as yea-saying, nay-saying and socially desirable responding) lead to over- or underreporting of innovative behavior.

---

Hans Baumgartner is Smeal Chair Professor of Marketing, Pennsylvania State University, USA (email: [hansbaumgartner@psu.edu](mailto:hansbaumgartner@psu.edu)). Bert Weijters is Associate Professor, Ghent University, Belgium (email: [bert.weijters@ugent.be](mailto:bert.weijters@ugent.be)).

The problem can also be illustrated using the specific example studied in the empirical part of this article. Assume that a researcher is interested in the attributions (in terms of locus, stability, and controllability) that consumers in different countries make about the cause-related marketing (CRM) activities of a global brand. If consumers attribute the activities to internal, stable, and controllable causes, positive outcomes (e.g., increased trust) will accrue to the firm or brand. A researcher may want to know how well the three attribution dimensions can be measured in different countries using a particular measurement instrument, whether the mean attributions differ across countries, or how strongly the attribution dimensions are correlated with each other and with other constructs. To conduct these cross-national comparisons with confidence, the researcher must know that the method of measurement used (self-reports based on different scale formats for the response scale) and the potential differences in response styles across countries (i.e., differences in how respondents in different countries use the response scale) do not confound the results. Unfortunately, researchers are often unsure about how to ascertain whether method variance is present and how method variance, if present, should be accounted for while conducting the necessary measurement invariance tests and substantive comparisons of interest.

In this article, we provide answers to these kinds of questions. In particular, we address two important issues related to common method variance in international marketing research. First, we review the most common procedures for detecting and controlling CMV and investigate whether international marketing researchers have shown concerns regarding common method variance in their studies and, if so, how they have tried to ascertain whether CMV was present and what methods they have used to account for CMV. To this end, we analyze survey-based *Journal of International Marketing* articles published between 2015 and 2019 and critically examine whether the authors of these articles used appropriate methods to deal with CMV. Second, we propose a procedure based on multisample structural equation modeling that researchers can use to model method effects in cross-cultural and cross-national studies. The recommended approach considers two different methods for dealing with CMV depending on whether method effects are directly measured or inferred from the substantive variables, and the procedure is illustrated using data from a 14-country study conducted in Europe.

## Statistical Approaches for Modeling Method Effects

Podsakoff et al. (2003) and Podsakoff, MacKenzie, and Podsakoff (2012) distinguish between procedural and statistical approaches to dealing with method effects. While we agree that researchers should make every effort to consider method effects when designing empirical studies, our focus here is on post hoc statistical remedies. Arguably the best-known and most widely used method for investigating CMV is the so-called Harman one-factor test. This “test” cannot be used to

control for CMV if it is present; its only purpose is to determine whether a researcher should be concerned about CMV. To implement the Harman test, a researcher must conduct an exploratory factor analysis on all the observed variables used in the analysis, and if the first factor is the only factor with an eigenvalue  $> 1$  and/or if the first factor accounts for  $> 50\%$  of the total variance in all items, CMV is said to be present. A closely related procedure, which is actually based on a statistical test, consists of estimating a one-factor confirmatory factor analysis model, and if this one-factor model is found to fit very poorly, the hypothesis that method variance is present is rejected. The Harman one-factor test has been severely criticized by many authors (e.g., Baumgartner, Weijters, and Pieters 2021; Hulland, Baumgartner, and Smith 2018; Podsakoff et al. 2003), and Hulland, Baumgartner, and Smith (2018, Appendix) forcefully argued that the test be put to rest. Similarly, Baumgartner, Weijters, and Pieters (2021, p. 234) recently concluded that “the Harman one-factor test is an ineffective tool for detecting CMV; researchers should stop using this likely misleading technique; and reviewers and editors should insist that it not be reported in published articles.” We will not repeat the many arguments against the Harman one-factor test here and instead refer the reader to the cited sources.

Many other methods for dealing with CMV have been suggested in the literature. Table 1 presents a classification of previous approaches similar to one proposed by Podsakoff et al. (2003). The rows of the table refer to techniques in which method effects are accounted for either at the scale level or at the item level. The columns refer to whether method effects are measured using variables other than the substantive items of interest or are inferred using the substantive variables themselves. In the former case, a measured method variable may assess a specific method effect directly (e.g., social desirability or acquiescence may be measured directly and used as control variables in the analysis), or a proxy measure (or so-called marker variable) aimed to capture an unspecified method is included in the analysis. In the latter case, method effects can be inferred from the substantive items either by including a latent method factor to model the nonsubstantive source(s) of covariation between the items (in addition to the substantive factors) or by specifying correlated uniquenesses (i.e., covariances between the unique factors) to model systematic sources of error between the items.

The two methods in the first column of Table 1 (i.e., cells A and B) can be used when a specific method effect is hypothesized to lead to CMV, which requires that explicit measures of the method effect in question be included in the questionnaire. For example, if a researcher believes that social desirability may confound the relationship between two substantive constructs, a social desirability scale could be included and used as a control variable. The control variable could be a single measure (with or without correction for unreliability) or a latent variable (measured by multiple indicators). If several sources of method effects are hypothesized, multiple control variables could be considered. The two methods in the first

**Table 1.** A Classification of Approaches for Controlling Method Effects.

		Are method effects measured or inferred?			
		Measured (Manifest)		Inferred (Latent)	
		Direct Measure	Proxy Measure	Method Factor(s)	Correlated Uniquenesses
<b>Are method effects controlled at the scale level or item level?</b>	Scale level	A. Partialing a directly measured method variable at the scale level (with or without control of measurement error, multiple directly measured method variables can be partialled)	C. Lindell and Whitney (2001) correlational marker variable technique applied at the scale level	E. Partialing an inferred but "measured" method factor (e.g., factor scores from an EFA à la Harman) at the scale level	n.a.
	Item level	B. Partialing a directly measured method variable at the item level (with or without control of measurement error, multiple directly measured method variables can be partialled)	D. Williams, Hartman, and Cavazotte (2010) Comprehensive Marker Variable Technique applied at the item level	F. Factor model with substantive factors and at least one method factor (although there could be several as in MTMM models)	G. Factor model with substantive factors and correlated uniquenesses instead of method factors

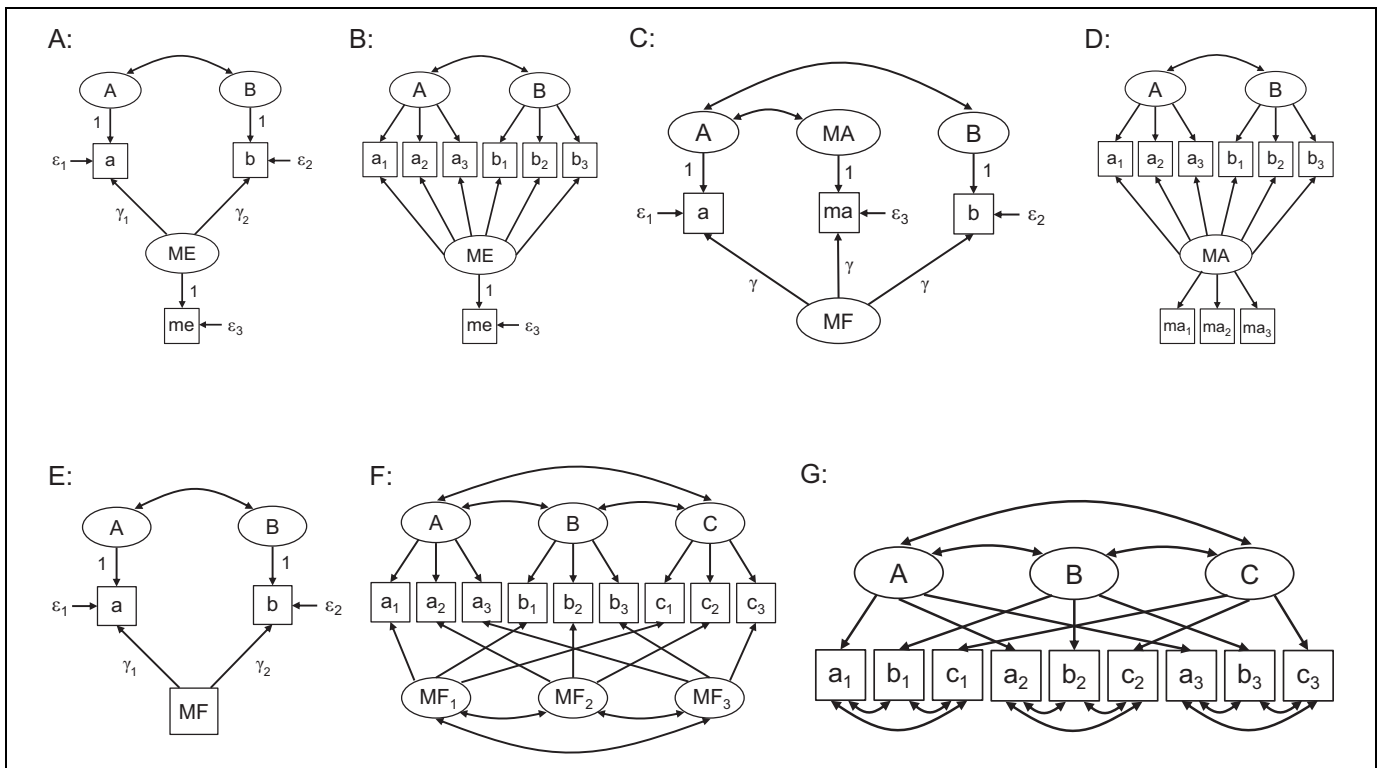
column of Table 1 differ depending on whether method variance is controlled for at the scale level (Table 1, top row; Figure 1, Panel A) or at the item level (Table 1, bottom row; Figure 1, Panel B). Podsakoff et al. (2003) recommend the latter because once method effects are purged from individual items, the construct of interest can be measured more accurately using the purified items.

It is important that the direct measure of method effects not be confounded with the substantive constructs. For example, suppose that a researcher wants to include social desirability as a control variable to account for potential method effects in the influence of attitudes and subjective norms on purchase intentions. It is likely that social norms and social desirability are substantively related, so controlling for social desirability will probably remove substantive variance from the social norm measures, leading to overcontrol of method effects (Steenkamp, De Jong, and Baumgartner 2010).

The two methods in the second column of Table 2 (i.e., cells C and D) are variations on the marker variable approach to accounting for method effects. The original marker variable method suggested by Lindell and Whitney (2001), sometimes called the correlational marker technique (Richardson, Simmering, and Sturman 2009), aims to remove CMV contamination at the scale level. Podsakoff, MacKenzie, and Podsakoff (2012) offer a long list of shortcomings of this technique, but the two major issues are the choice of appropriate marker variables and the restrictive assumptions underlying the method. Lindell and Whitney (2001) argue that marker variables should be highly reliable (which generally implies that a multi-item scale be used) and that they should be theoretically unrelated to at least one of the substantive variables (i.e., the true correlation between the marker variable and one of the substantive constructs should be zero). Ideally, a questionnaire should include such a marker variable

by design, but Lindell and Whitney suggest that researchers may also choose the smallest correlation among the substantive correlations as an estimate of CMV. Although the requirement that a marker variable be substantively unrelated to at least one substantive variable is a necessary condition for accurately capturing method effects, it is not sufficient, as a marker variable also must tap some type of method effect to be useful for removing method variance. Even more problematically, the ad hoc choice of the smallest correlation among the substantive correlations as an approximation of CMV makes little conceptual sense because it is unclear what the smallest correlation represents (substantive variance, method variance, other systematic biases besides method variance, random deviations from zero, etc.). Furthermore, using the smallest correlation makes it a priori unlikely that evidence of CMV will be detected because the smallest correlation is often close to zero. Unfortunately, as we will show, researchers have mostly used nonsensical marker variables when using the marker variable approach.

The marker variable technique also imposes rather restrictive assumptions on the control of method variance. Although Lindell and Whitney (2001) derive their method on the basis of partial correlations, the underlying model can be depicted as a structural equation model as shown in Figure 1, Panel C. As already discussed, one of the substantive variables (Construct B in Figure 1, Panel C) must be conceptually unrelated to the marker variable, and this is shown by the missing bidirectional arrow between B and MA in the figure. The only open path between B and MA is through the underlying method factor (MF). Because the model in Figure 1, Panel C, is exactly identified (i.e., it has zero degrees of freedom), this assumption cannot be tested. Furthermore, it is apparent that the loadings of a, b, and ma on the method factor are identical, meaning that the hypothesized method factor influences all three variables



**Figure 1.** Method factor models.

**Notes for Panel A:** A and B are substantive constructs, and a and b are the observed measures of these constructs (usually an average of multiple observed measures of a given construct). The covariance between A and B is the parameter of primary interest. ME is a hypothesized method effect, and me is the observed measure of the method effect. To correct for measurement error in the observed measures, the variances of  $\varepsilon_1, \varepsilon_2$ , and  $\varepsilon_3$  are set to the variance of the observed measure minus the reliability of the observed measure. Alternatively, a factor model with multiple indicators could be specified for ME. If no correction for unreliability is applied, the variances of  $\varepsilon_1, \varepsilon_2$ , and  $\varepsilon_3$  are set to zero and  $A = a, B = b$ , and  $ME = me$ .

**Notes for Panel B:** A and B are substantive constructs, and  $a_1$  to  $a_3$  and  $b_1$  to  $b_3$  are the observed measures of constructs A and B, respectively. The effects of ME on the indicators of A and B are freely estimated, but the parameters are not shown explicitly in the figure. Error terms associated with  $a_1$  to  $a_3$  and  $b_1$  to  $b_3$  are omitted for simplicity. See Panel A for additional details.

**Notes for Panel C:** A and B are substantive constructs, and a and b are the observed measures of these constructs (usually an average of multiple observed measures of a given construct). The covariance between A and B is the parameter of primary interest. MA is the marker construct, and ma is the indicator of the marker construct. MF is a latent method factor. For identification, the covariance between MA and B (or A) must be fixed to zero. To correct for measurement error in the observed measures, the variances of  $\varepsilon_1, \varepsilon_2$ , and  $\varepsilon_3$  are set to the variance of the observed measure minus the reliability of the observed measure. If no correction for unreliability is applied, the variances of  $\varepsilon_1, \varepsilon_2$ , and  $\varepsilon_3$  are set to zero and  $A = a, B = b$ , and  $MA = ma$ .

**Notes for Panel D:** This model looks similar to the model in Panel B, but it is quite different. A and B are substantive constructs, and  $a_1$  to  $a_3$  and  $b_1$  to  $b_3$  are the observed measures of constructs A and B, respectively. MA is a marker variable construct, and  $ma_1$  to  $ma_3$  are the indicators of the marker construct. The covariances of A and B with MA are constrained to zero, but the loadings of  $ma_1$  to  $ma_3$  on MA are constrained to be equal to the values of these loadings in a congeneric three-factor model in which the covariances between A, B, and MA are freely estimated. The effects of MA on the indicators of A and B are either constrained to be equal (in the so-called Method-C model) or freely estimated (in the so-called Method-U model). Model comparison tests should be conducted to evaluate which model is preferable. The covariance between A and B in the model with method effects is also compared to the covariance between A and B in the model without method effects to evaluate whether the inclusion of method effects changes the substantive conclusions. Error terms associated with  $a_1$  to  $a_3, b_1$  to  $b_3$ , and  $ma_1$  to  $ma_3$  are omitted for simplicity.

**Notes for Panel E:** A and B are substantive constructs, and a and b are the observed measures of these constructs (usually an average of multiple observed measures of a given construct). The covariance between A and B is the parameter of primary interest. MF is an inferred but "measured" method factor that is based on the factor scores estimated from a one-factor model based on all individual measures of A and B (e.g.,  $a_1$  to  $a_3$  and  $b_1$  to  $b_3$ ).

**Notes for Panel F:** A, B, and C are substantive constructs, and  $a_1$  to  $a_3, b_1$  to  $b_3$ , and  $c_1$  to  $c_3$  are the observed measures of constructs A, B, and C, respectively.  $MF_1, MF_2$ , and  $MF_3$  are three inferred method factors corresponding to three types of items. Other method factor specifications (including a model with a single inferred method factor) are possible. All loadings and the covariances between the method factors are freely estimated in the model shown, but simpler specifications (e.g., a model with uncorrelated method factors) may be considered. Error terms associated with the observed measures are omitted for simplicity.

**Notes for Panel G:** A, B, and C are substantive constructs, and  $a_1$  to  $a_3, b_1$  to  $b_3$ , and  $c_1$  to  $c_3$  are the observed measures of constructs A, B, and C, respectively. Instead of method factors, correlated uniquenesses are specified to model the method effects corresponding to each type of item. With three indicators per substantive factor, the model in Panel G is identical to the model in Panel F with uncorrelated method factors. Error terms associated with the observed measures are omitted for simplicity.

equally. Again, this assumption cannot be tested. The model in Panel C allows a direct path between A and MA, and if this path is constrained to zero, the restrictive assumptions contained in the model could be tested. Most likely, this model would not fit

the data well. Constraining the loadings of the method factor on the substantive indicators (a and b) to be equal is often considered too restrictive (Podsakoff et al. 2003). But the truly unrealistic assumption is that the effects of the method factor on the

**Table 2.** Dictionary Used for the Systematic Review.

Concept	Search Terms
Survey	"survey," "questionnaire," "self_report," "Likert"
CMV	"common_method," "method_variance," "method_bias," "CMV"
Harman	"Harman," "Harmon," "Harman*," "single_factor_test"
Partial correlation	"Lindell," "marker," "Lindell_and_Whitney," "marker_variable," "marker_indicator," "partial_correlation technique," "partial_correlation_procedure"
Method factor	"method factor," "method_factor," "common_latent_factor," "unmeasured_latent"

substantive items are as strong as the effect of the method factor on the marker variable, which is supposed to be a direct proxy for method effects. Overall, because of the many problems afflicting the Lindell and Whitney marker variable technique, we are skeptical about its value as a control for method effects (consistent with Podsakoff et al. [2003]). The popularity of the technique is probably due to its ease of use and the fact that the marker variable method usually leads to the desired finding that method variance is not a problem.

Williams, Hartman, and Cavazotte (2010) propose a more sophisticated version of the correlational marker variable technique that enables a correction for method effects at the item level. The technique, known as the "Comprehensive CFA Marker Technique" (see Figure 1, Panel D) extends earlier applications of confirmatory factor analysis designed to control for method variance through marker variables (referred to as the CFA Marker Technique by Richardson, Simmering, and Sturman [2009]). This rarely used procedure is complex: (1) the proposed factor model includes both substantive factors and a latent marker factor measured by multiple indicators; (2) the latent marker factor is specified to be uncorrelated with the substantive factors, although the loadings of the marker indicators on the latent marker factors are fixed to the values obtained from a CFA model in which these factor correlations are freely estimated; (3) the latent marker factor can have equal or unequal effects on the indicators of the substantive factors; and (4) the correlations between the substantive factors are explicitly compared to the model with and without method effects.

Despite certain advantages (e.g., control of method effects at the item level, correction for measurement error in the indicators of the marker factor, consideration of equal vs. unequal method effects, examination of the influence of method effects on the substantive correlations), the Comprehensive CFA Marker Technique fails to resolve the basic problem of the marker variable technique: How can a marker variable serve as a measured proxy for method effects when the concept of method variance remains undefined? Williams, Hartman, and Cavazotte (2010, p. 507) are aware of this conundrum and recommend that "the definition of a marker variable [should] be expanded so that not only is a marker variable defined as a variable that is not expected to be theoretically related to substantive variables in the model but . . . is also defined as capturing or tapping into one or more of the sources of bias that can occur in the measurement context for given substantive

variables being examined." Still, the authors also explicitly differentiate marker variables from measured method effect variables. It may be tempting to hope that one can find a marker variable that miraculously captures all the method effects that impinge on the observed measures in a survey, but this is likely futile and reflects "deus ex machina" thinking. It seems more realistic to try to identify likely method effects a priori and measure these method effects directly, or to infer method effects implicitly as discussed next.

The third column in Table 1 (i.e., cells E and F) refers to methods in which method effects are accounted for using an inferred latent method factor (or possibly several inferred method factors). When method effects are modeled at the scale level, a single-factor exploratory or confirmatory factor analysis is conducted first, then factor scores for this factor (or possibly averages of all items) are computed, and finally these factor scores are used as a control variable similar to the partial correlation approach at the scale level with a measured method variable (see Figure 1, Panel E). Like the Harman one-factor test, this method incorrectly assumes that the first factor is a "pure" method factor. Particularly in the (rather common) situation in which all substantive variables in the analysis are positively related, the first factor will likely capture primarily substantive variance, which implies that the presumed method factor will actually remove substantive variance. The method has several other weaknesses (e.g., method effects are only controlled at the scale level, the approach is piecemeal and requires the potentially problematic computation of factor scores) and should therefore not be used to account for method effects.

A preferred approach to modeling method effects via an inferred method factor (or several method factors) is to assume that the observed measures are a function of both a substantive factor and a method factor (in addition to a unique factor). In the simplest such model, a single method factor influences all observed measures. At least two substantive factors are needed for this model to be identified, and often the method loadings have to be constrained to be equal to obtain a converged solution. This model is preferable to the corresponding scale-level model because the substantive factors are modeled explicitly (which avoids the confounding of substantive and method variance to some extent), but in general it is not clear what the method factor represents. When each of several constructs of interest is measured via multiple methods of measurement (e.g., three different response scale formats are used to measure

several constructs), the resulting model is called the multitrait-multimethod (MTMM) model in the literature (see Figure 1, Panel F). In principle, this is an attractive model for representing method effects (in part because method effects can be inferred from the substantive measures themselves), and the model has been used in numerous studies; however, in practice there are several weaknesses (see Baumgartner and Weijters [2019] for a recent discussion). First, researchers often use methods that are quite similar, and if the method factors are allowed to be correlated, it is not clear whether the presumed method factors really capture method effects (Marsh 1989). Second, simulations show that inferred method factors may fail to yield accurate estimates of method effects (Richardson, Simmering, and Sturman 2009). Third, MTMM models are notoriously difficult to estimate, and researchers sometimes simplify the model by imposing possibly unrealistic assumptions, such as zero correlations between the method factors or equality of method loadings, to avoid improper solutions and nonconvergence issues.

Our recommendation is to use the model based on inferred method factors only under specific circumstances. For example, assume that two constructs are measured with balanced scales, where half the items in each scale are reverse-keyed. If the reversed items have not been recoded, the regular items should have a positive loading on the substantive constructs, and the reversed items should have a negative loading. A method factor on which all the items of both scales have a positive loading will capture response tendencies such as acquiescence that indicate a lack of sensitivity to the keying direction of the items (Weijters, Baumgartner, and Schillewaert 2013). In this case, the interpretation of the method factor has at least face validity because somebody who agrees with both an item and its opposite probably does not respond on the basis of substantive considerations.

The final model for representing method effects is similar to the previous model, but instead of specifying a method factor, so-called correlated uniquenesses (i.e., covariances between the unique factors or error terms) are used to account for method effects (see Figure 1, Panel G). The correlated uniqueness (CU) model was originally promoted as a way of avoiding the estimation problems associated with MTMM models. Another advantage is that the CU model does not assume unidimensionality of method effects. However, the CU model (1) is less parsimonious when there are many indicators, (2) makes it impossible to specify correlated method effects across different methods, and (3) does not allow for the study of antecedents, correlates, or consequences of method effects. In general, we prefer method factor models to CU models, although with three indicators per factor, the CU model is usually identical to a method factor model with uncorrelated method factors.

In summary, although many approaches have been suggested to account for method effects in empirical data, only partialing a directly measured method variable at the item level and confirmatory factor analysis using an inferred method factor (or several method factors) can be recommended for future

use. We next report a study of *Journal of International Marketing* articles to examine how international marketing researchers have dealt with method effects in their research.

## Study I

To investigate how CMV is dealt with in *Journal of International Marketing* research, we performed a systematic review of articles published in the journal during the last five years (i.e., 2015–2019). The goal was to assess how often authors discuss CMV, how they test for its presence and potential impact on the results, and how they attempt to control for CMV when they deem it necessary to take corrective action.

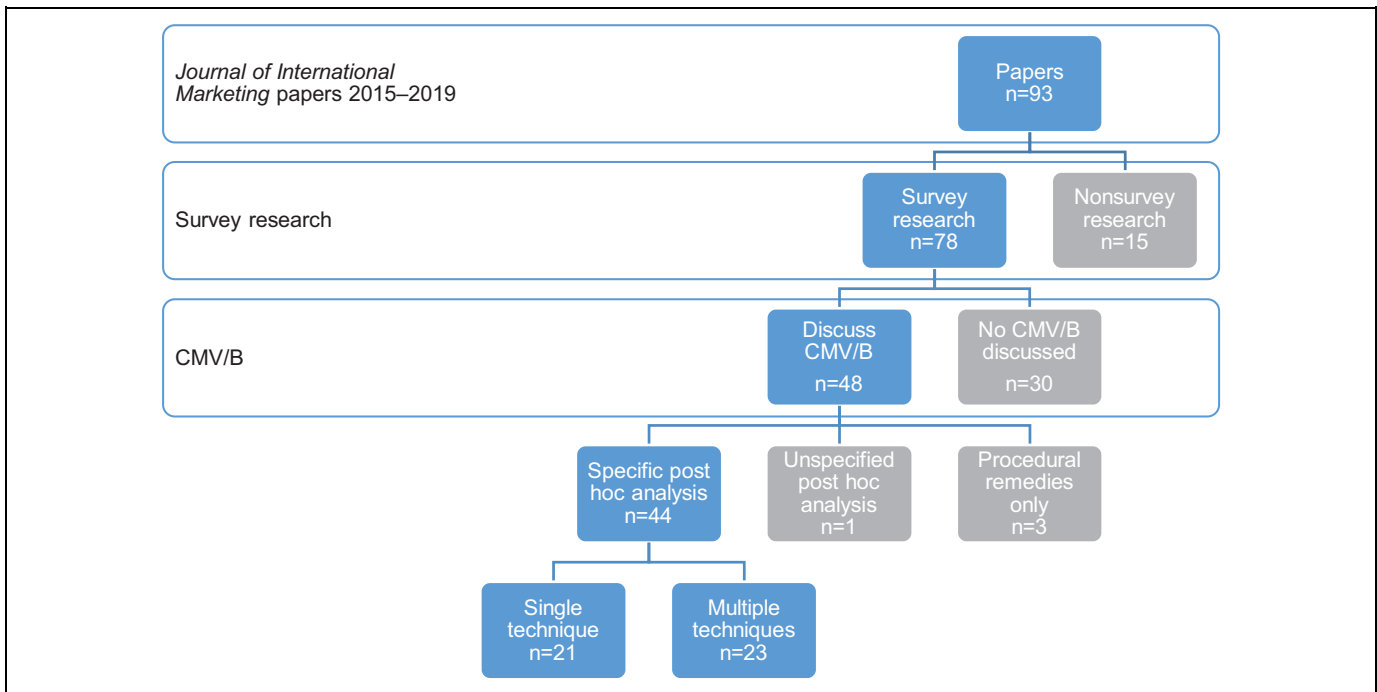
## Method

We analyzed all *Journal of International Marketing* articles published in the period from January 2015 to December 2019 ( $N = 93$ ) using the Quanteda package in R. The files were assigned to a corpus and tokenized. We then searched the documents for instances of terms related to CMV using the search terms specified in the dictionary in Table 2, which includes five categories of terms: (1) Category 1 aims to identify articles based on survey data; (2) Category 2 tries to identify articles that discuss CMV; (3) Category 3 seeks to identify articles in which researchers used the Harman one-factor test; (4) Category 4 attempts to identify articles in which researchers used partial correlation procedures based on measured method effects (including marker variables) to correct for method effects; and (5) Category 5 tries to identify articles in which researchers inferred method effects on the basis of a latent method factor.

We then read the paragraph(s) of the articles flagged as discussing CMV and eliminated false positives (e.g., a reference to an author named Harmancioglu). We also classified articles according to whether they used the Harman one-factor test or any of the other approaches listed in Table 1 to check for CMV. Finally, we made some additional notes on the conclusions drawn by the authors based on their CMV analyses.

## Results

Figure 2 shows a tree diagram of the findings based on the Quanteda analysis (using the dictionary in Table 1) and the corrections applied after reading the relevant paragraphs dealing with CMV in each article. Of the 93 *Journal of International Marketing* articles, 78 (or 84%) were survey-based. Most of the survey-based articles (48 of 78, or 62%) discussed CMV. Of the 48 survey-based articles that offered a discussion of CMV, 3 articles did not report any post hoc CMV analyses, and the authors of one article only stated that they “conducted common method bias checks.” The 3 articles that did not report post hoc tests did mention effective procedural remedies, however (e.g., use of multisource data or two key informants). Of the 44 articles in which specific post hoc tests were conducted,



**Figure 2.** Key results of the systematic review.

**Table 3.** Contingency Table of Observed Post Hoc CMV Techniques in Study 1.

	1	2	3	4	5	6
1. Harman one-factor test	23	0	2	12	1	6
2. Partialing a measured method at the scale level (Table 1, Cell A)	0	1	0	1	0	0
3. Partialing a measured method at the item level (Table 1, Cell B)	2	0	2	1	0	0
4. Lindell & Whitney marker technique (Table 1, Cell C)	12	1	1	27	0	3
5. Comprehensive Marker Variable Technique (Table 1, Cell D)	1	0	0	0	2	1
6. Inferred latent method factor(s) (Table 1, Cell F)	6	0	0	3	1	14

Note: The cells in this table display the frequency of occurrence (diagonal) and co-occurrence (nondiagonal) of the various techniques.

21 used a single method and 23 used two or more. As reported in Table 3, the most commonly used approaches were the Lindell and Whitney marker technique (27) and the Harman one-factor test (23), and these techniques were also sometimes reported as being used together (12). Note that we treated articles in which the authors performed a single-factor confirmatory factor analysis and interpreted a poorly fitting one-factor model as evidence that no method variance was present as applications of the Harman one-factor test, because both are based on the same (faulty) logic. Factor models with substantive factors and at least one inferred latent method factor were also reported somewhat frequently (14), sometimes in follow-up analyses to a one-factor test (6). The methods based on partialing a directly measured method variable at either the scale level (1) or the item level (2) and versions of the William, Hartman, and Cavazotte (2010) Comprehensive Marker Variable Technique (2) were used infrequently, and partialing an inferred method factor at the scale level and factor models with correlated uniquenesses were not reported in our sample of articles.

Closer inspection of these articles yields some further insights. Most articles dedicated at least a full paragraph to discussing CMV. But strikingly, only in one article did the authors conclude that the analysis indicated a potential risk of CMV bias, leading the authors to statistically control for CMV in their main analysis. The other articles consistently concluded that CMV did not pose a threat to validity (e.g., “we can safely conclude that CMV does not seem to be a problem in this study”; “taken together, the results of all of the tests indicate that common method bias does not seem to be a threat within our data [though it cannot be ruled out completely] and is unlikely to explain any of the results of our hypothesis tests”). In some instances, this conclusion may be plausible, but in many cases, there are no valid grounds for such reassurances, as we discuss next.

## Discussion

Most survey-based *Journal of International Marketing* articles that have recently been published discuss the problem of CMV

and attempt to do something about it. This is a reassuring finding, but unfortunately the methods that are used to detect and control for CMV are often suboptimal.

First, it is surprising that the Harman one-factor test is still commonly used and reported despite the criticism leveled at the test in the influential review article by Podsakoff et al. (2003) almost 20 years ago. Ironically, many articles cite Podsakoff et al. but ignore their recommendations. As stated more recently by Hult, Baumgartner, and Smith (2018, p. 106), "the Harman test is entirely nondiagnostic about the presence of common method variance in data. Researchers should stop going through the motions of conducting a Harman test and pretending that they are performing a meaningful investigation of systematic errors of measurement." On the positive side, the Harman one-factor test is rarely reported as the only test of method variance, as 18 of the 23 articles in which it is conducted also report another CMV test.

Second, the use of marker variables is remarkably popular, but too many problems occur with this approach. First, some authors do not report which marker variable they used, which makes it hard to evaluate how relevant the marker is or how the results could be replicated. Second, some marker variables cannot reasonably be expected to be valid indicators of common method variance in Likert-type rating scales. Examples of questionable marker variables include the number of years a firm has been purchasing from abroad, occupation (with the question wording and response format left unspecified), the respondent's position within the company (not further specified), or demographics such as age, gender, income, and education. It is not clear how such marker variables could meaningfully capture typical sources of common method variance such as impression management, consistency bias, or response styles. Many authors mention that the marker variable must be conceptually unrelated to at least one focal construct, but they do not seem to realize that this is not a sufficient condition because the marker variable must also serve as a proxy variable for any method effects that might confound the results.

Third, the use of an inferred method factor can be useful, but several instances of problematic practices emerged here as well. For example, some researchers simply added a method factor to a model already containing the substantive factors (i.e., each substantive indicator loads on both a substantive factor and a method factor) and concluded that because the method loadings were small or at least smaller than the substantive loadings, CMV was not a problem. Obviously, concluding that there is more substantive variance than method variance is not the same as concluding that there is no method variance. Just because method variance is smaller than substantive variance does not necessarily mean that method variance can be ignored in subsequent analyses. Also, if all items load on a common method factor, it is only possible to distinguish method variance from substantive variance under specific circumstances. For example, as discussed earlier, if a scale contains an equal number of regular and reversed items, a substantive construct can be distinguished from a method factor; however, when all substantive constructs are positively

correlated and the substantive scales contain no reverse-keyed items, substance and method are likely confounded.

To summarize, researchers are still using the Harman one-factor test despite clear evidence that it cannot validly detect the presence and severity of method variance. There are also serious problems with the concept of a marker variable, and most marker variables used in practice lead to invalid corrections for CMV. The preferred methods are based on using a directly measured method effect or an inferred method factor to account for method variance at the item level, and the latter method is only useful under certain circumstances. We next discuss how these two methods can be used in international marketing research, and we illustrate the proposed approach with a cross-national study involving 14 European countries.

## **How to Model Common Method Variance in International Marketing Research**

There are two issues that researchers must address when trying to account for common method variance in international marketing research. First, CMV must be captured using an approach that is conceptually meaningful. This excludes approaches such as the Harman one-factor test and most applications of the marker variable technique. Second, CMV must be modeled in a principled way in all countries that are being compared. This is in contrast to the usual practice of ignoring method effects in the main analysis because CMV is presumably not a threat (based on evidence obtained via questionable post hoc CMV approaches).

With respect to the first issue, the best way to deal with CMV is to carefully evaluate the procedural remedies discussed in such sources as MacKenzie and Podsakoff (2012). Although our focus is on post hoc statistical approaches, an a priori consideration of potential sources of method effects may point to specific threats that should be addressed head-on in the empirical study. For example, if a research team expected social desirability to influence respondents' answers to the substantive questions, they should include a recommended social desirability scale in the study (e.g., Steenkamp, De Jong, and Baumgartner 2010). Often, it will be difficult to identify potential threats a priori, and there may be too many potential threats to include separate measures for all of them. However, an excellent approach for dealing with method effects that are quite common in international marketing research is to include a scale consisting of items that are (1) conceptually unrelated to each other and (2) conceptually unrelated to the substantive constructs studied. In other words, the items in this scale should be heterogeneous in content (i.e., because the items are unrelated and thus lack common content, researchers can use them to construct a "pure" measure of method effects) and unrelated to the substantive constructs in the study. Using the items in this scale (which we call a response style scale), the researcher can then calculate various indices of stylistic responding, which are often sources of method effects and which have been shown to vary systematically across cultures and countries (Baumgartner and Weijters 2015). Among the most common



response styles are (dis)acquiescent and net acquiescent responding, extreme responding, and midpoint responding, and they can be computed for each person from the items in the response style scale as follows (Weijters, Schillewaert, and Geuens 2008):

1. Acquiescent responding:  $ARS = [f(4) \times 1 + f(5) \times 2] / k$
2. Disacquiescent responding:  $DARS = [f(2) \times 1 + f(1) \times 2] / k$
3. Net acquiescent responding:  $NARS = [f(1) \times 1 + f(2) \times 2 + f(3) \times 3 + f(4) \times 4 + f(5) \times 5] / k$
4. Extreme responding:  $ERS = [f(1) \times 1 + f(5) \times 1] / k$
5. Midpoint responding:  $MRS = [f(3) \times 1] / k$ ,

where  $f(x)$  refers to the frequency with which response option  $x$  is chosen by a respondent across all  $k$  items in the response style scale. In these expressions, we have assumed that there are five response options (because five-point scales seem to be most common in practice), but extensions to more than five response categories are straightforward. Prior research has shown, for example, that respondents from individualistic cultures are more likely to engage in extreme responding and less likely to engage in (net) acquiescent and midpoint responding, whereas the opposite is true for respondents from collectivist cultures (see the review in Baumgartner and Weijters 2015). Because these response styles vary systematically across cultures, it is important to take these differences into account; otherwise stylistic response differences could be mistaken for substantive differences.

Instead of directly measuring method effects, a researcher may sometimes specify a method factor meant to infer method effects from the substantive items. Although this approach has the advantage that it does not require the inclusion of separate response style items, it must be used with great care. For example, specifying a general method factor on which all the items are allowed to load (in addition to the substantive factors) is unlikely to yield a valid test of method effects in general. As discussed earlier, one exception occurs when the substantive constructs are measured by balanced scales (i.e., scales in which half the items are keyed in one direction and half in the opposite direction), in which case the loadings of the reverse-keyed items on the substantive factor will be negative (if reversed items have not been recoded) and the loadings on the method factor will be positive. Another option is to design the study in such a way that method effects can be distinguished from the substantive effects. For example, a researcher can measure each substantive factor with several different methods (e.g., different scale formats) and then specify an MTMM model in which the variance in each measure can be partitioned into substantive variance, method variance, and unique variance. Unfortunately, it is often difficult to estimate such models (particularly when the methods are conceptually related), and if the methods are allowed to be correlated, there is no guarantee that the presumed method factors actually measure method effects (for examples and further discussion, see Marsh [1989] and Baumgartner and Weijters [2019]).

With respect to the second issue, models based on data from several cultures or countries can be specified using a multi-sample analysis (fixed-effect specification) or a hierarchical analysis (random-effect specification). Because the number of cultures or countries for which data are available is usually relatively small, we focus on the former. We recommend that international marketing researchers specify a series of models and conduct explicit model comparisons to evaluate whether method effects are present and need to be accounted for in testing the substantive hypotheses of interest. In the following discussion, we assume that the model of interest is a factor model in which the relationships between the substantive constructs are modeled as covariances. However, a model with directed relationships between the constructs could also be specified.

The first model is a factor model in which only the substantive factors are considered and all parameters are freely estimated in all cultures or countries. This model serves as a baseline model. Next, methods effects are introduced either by (1) relating each observed measure to a directly measured method effect or (2) specifying a latent method factor (or multiple method factors) in an effort to infer method effects from the substantive items. Method effects are present if (1) the direct measure of method effects has a significant influence on the observed measures (an overall test can be conducted by comparing the model in which all method effects are constrained to zero with a model in which the method effects are freely estimated) or (2) the loadings of the substantive items on the inferred method factor(s) are significant (an overall test can be conducted by comparing the model with method effects to the baseline model). If a researcher finds method effects to be present, they may consider further models in which the method effects or method loadings are constrained to be equal across items. A comparison of this model with the previous model tests whether method effects are uniform across items. Finally, the preferred model specification based on these model comparisons can be compared across cultures or countries. Unless a researcher wants to compare method effects or relationships between method effects and other constructs across cultures or countries, it is not necessary to establish the (metric or scalar) invariance of method effects across groups. However, researchers often want to compare scores on substantive constructs (e.g., means) or relationships between substantive constructs across cultures or countries. If such comparisons are to be conducted, it is necessary to establish the metric or scalar invariance of the purified observed measures of the substantive constructs (i.e., the observed measures that have been purged of method effects) (Steenkamp and Baumgartner 1998). If a researcher only wants to compare covariances or structural relations across groups, it is sufficient to assess the metric invariance of the substantive loadings. In contrast, when construct means are to be compared across cultures or countries, the scalar invariance of the items must be tested as well (for details, see Steenkamp and Baumgartner 1998). Next, we provide additional details in the context of an empirical application.

## Study 2

The goal of this study is to provide an empirical illustration of the previously described procedure for assessing and controlling CMV across countries. The empirical example in this study focuses on consumer perceptions of a global brand engaging in cross-national cause-related marketing (CRM), that is, “activities by firms to contribute to designated causes” (Varadarajan and Menon 1988, p. 60). In line with previous research in this domain (Klein and Dawar 2004; Sen and Bhattacharya 2001), we applied Weiner’s attribution model (Weiner 1980, 1985; Weiner, Graham, and Chandler 1982) to estimate individual variation in consumer perceptions of a brand’s CRM initiatives. Weiner’s model distinguishes three dimensions of attributions that influence the inferences consumers draw from an event or behavior, such as a specific CRM initiative: (1) the locus of the behavior (the event that triggers the CRM initiative), which can be internal or external to the company; (2) the stability of the behavior, which can be long-term or short-term; and (3) the controllability of the behavior, which can be within or outside the control of the company. These three attribution factors are relevant because positive outcomes for the firm (such as increased trust) will mostly occur when a firm’s CRM initiatives are attributed to internal, stable, and controllable causes (Ellen, Webb, and Mohr 2006).

For our illustration, we used the three attribution dimensions as our substantive factors. We measured each factor using three different methods (i.e., different response scales), which enabled us to specify a model with several latent (inferred) method factors. In addition, we considered a single directly measured method effect, which is described subsequently. Because we had data from 14 different countries, we can demonstrate how CMV can be dealt with in a relatively complex cross-national study, and we can also show how two recommended methods can be used to test and account for CMV in international marketing research.

## Method

The data were collected through an online survey among members of a market research online panel in 14 countries (total  $N = 11,970$ , with quotas for gender and age category). The number of respondents per country is as follows: Belgium, 899; Denmark, 838; France, 893; Germany, 881; Italy, 836; Netherlands, 927; Poland, 708; Romania, 866; Slovakia, 941; Spain, 831; Sweden, 871; Switzerland, 858; Turkey, 813; and UK, 808. The questionnaire was translated by professional translators using the back-translation procedure. Respondents read a description of plausible but fictitious CRM initiatives by a global soft drink brand. We used different scenarios to vary locus, stability, and controllability, but each respondent was randomly assigned to a single scenario. For our purposes, the different scenarios simply created variance in the attribution measures, which are the focus of our study. An illustrative scenario is as follows: “Children are the building blocks of our society. Starting from that conviction the following initiative

was developed. In November [the brand] will give away a fixed amount of money for every can that consumers buy. The aid is for the benefit of a European organization that supports children during their rehabilitation after a traffic accident.” Appendix A lists all eight scenarios used in the study.

The questionnaire consisted of four major sections. First, respondents read a scenario that briefly described a hypothetical CRM initiative by a global soft drink brand. Second, respondents rated the CRM initiative on nine attribution measures. Third, respondents answered 16 questions about a heterogeneous set of issues unrelated to the CRM initiative. Fourth, respondents provided some background information, which is not of interest here.

We measured the three attribution dimensions (locus, stability, and control) using three items each. To make the questions less repetitive, the questions asking about each attribution dimension were varied slightly, but the major difference was that three different response scales were used for each dimension. The attribution questions and the scale formats are reported in Table 4.

The scale used for measuring method effects directly consisted of 16 items that were deliberately heterogeneous in content (Greenleaf 1992). Examples include “I am a homebody”; “a college education is very important for success in today’s world”; “when I see a full ashtray or wastebasket, I want it emptied immediately”; “I eat more than I should”; and “no matter how fast our income goes up, we never seem to get ahead.” The heterogeneous content of the items makes them well suited for quantifying differences in nonsubstantive response patterns, and the scale has been used for this purpose in cross-national research (Weijters, Baumgartner, and Geuens 2016). Respondents indicated their answers to the Greenleaf items on a seven-point Likert scale ranging from “strongly disagree” to “strongly agree.” Although several different sources of method effects could be computed from the Greenleaf scale (ARS, DARS, ERS, and MRS, as described earlier), we only computed NARS (which is equivalent to the average score across the 16 items). NARS reflects systematic scale usage differences in net endorsement of items regardless of content (Baumgartner and Steenkamp 2001).

## Results

The data analysis in this case must account for the following complexities. First, we have data from 14 different countries. We use multisample structural equation modeling to represent the data, starting with a series of models in which no invariance constraints are imposed across countries (Models M1–M8) and eventually testing (partial) metric invariance of the substantive loadings (Models M9 and M10; see Table 5). Second, we and demonstrate two different approaches for modeling method effects. One is based on a measured method effect (net acquiescence response style, or NARS, which reflects scale usage differences regardless of the substantive content of the items). The other is an inferred method factor that takes into account the fact that the same response scale is used to measure

**Table 4.** Questions and Scale Formats Used in Study 2.

	Scale Format A	Scale Format B	Scale Format C
	Very unlikely Unlikely Somewhat unlikely Neutral Somewhat likely Likely Very likely	Open question with numeric field accepting scores from 0 to 100	1 = “not at all” to 7 = “very strongly”
<b>Factor:</b> Locus	How likely is it that each of the following parties is a source of the initiative? Consumers of [the brand], Personnel of [the brand], [the brand] itself, Others	Please divide 100 points between the different parties as an indication of how responsible you feel that each party is for the initiative. Consumers of [the brand], Personnel of [the brand], [the brand] itself, Others	To what extent do you believe that [the brand] is responsible for the initiative described earlier?
Stability	This type of initiative represents something stable and ongoing for [the brand]	To what extent do you believe that [the brand]’s support for this kind of initiative is something stable and ongoing? Please give a rating from 0 through 100, where 0 means “this is a one-time initiative for [the brand]” and 100 means “this is an enduring, stable commitment for [the brand].”	To what extent do you believe that [the brand]’s support for this kind of initiative is something stable and ongoing?
Controllability	How likely is it that [the brand] has control over this sort of initiative?	Please indicate to what extent you believe that [the brand] is in control of the initiative on a scale from 0 through 100, where 0 means [the brand] is not in control and 100 means [the brand] has full control.	To what extent is [the brand] in control of the initiative?

Note: For locus, we used the score for “[the brand]” itself (the scores for other parties involved were not included in the analysis). The second item of locus (locus-b) uses a constant sum format. The scores for all items in Column B were divided by 15 to make the scale ranges similar to those of the other items.

**Table 5.** Model Fit Comparison.

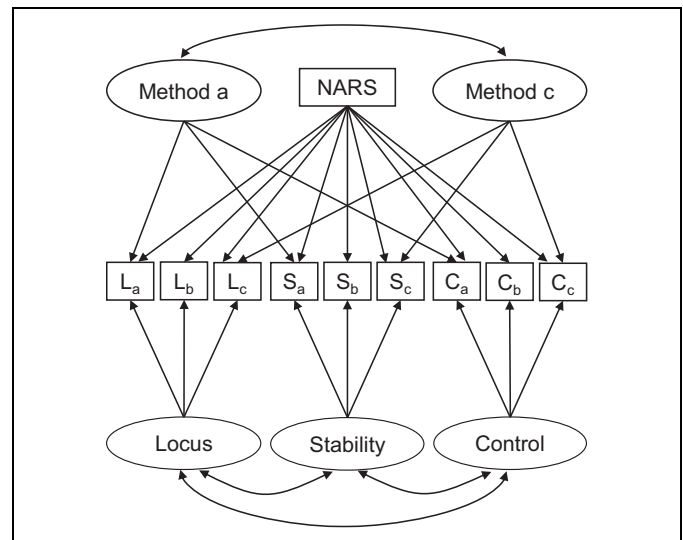
	$\chi^2$	DF	RMSEA	CFI	TLI	SRMR	BIC
Models with no invariance constraints across countries							
M1: Baseline model with three traits and no effects of NARS on items	4,294.05	462	.098	.953	.935	.055	360,899.94
M2: Model with three traits and unequal effects of NARS on items	3,849.32	336	.111	.956	.918	.040	361,638.37
M3a: Model with three traits and equal effects of NARS on items	4,113.90	448	.098	.955	.936	.044	360,851.25
M3b: Model with three traits and equal effects of NARS on items (except one item)	3,977.33	434	.098	.956	.936	.042	360,846.15
M4: Model with three traits, two correlated method factors, and no effects of NARS on items	2,092.47	364	.075	.979	.963	.052	359,618.60
M5: Model with three traits, two uncorrelated method factors, and no effects of NARS on items	2,870.98	378	.088	.969	.949	.053	360,265.64
M6: Model with three traits, two correlated method factors (equal method loadings within each method factor), and no effects of NARS on items	2,196.79	420	.070	.978	.967	.052	359,197.07
M7: Model with three traits, two uncorrelated method factors (equal method loadings within each method factor), and no effects of NARS on items	2,985.45	434	.083	.968	.954	.054	359,854.27
M8: Model with three traits, two correlated method factors (equal method loadings within each method factor), and equal effects of NARS on items (except one item)	1,893.06	392	.067	.981	.970	.038	359,156.27
Models with invariance constraints across countries							
M9: M8 with invariant substantive (trait) loadings across countries	2,316.66	470	.068	.977	.969	.051	358,847.43
M10: M9 with partially invariant substantive (trait) loadings across countries	2,146.01	463	<b>.065</b>	<b>.979</b>	<b>.972</b>	<b>.045</b>	<b>358,742.51</b>

Note: The model with the lowest (for badness-of-fit measures) and highest (for goodness-of-fit measures) fit measures that take model parsimony into account are shown in boldface. Model M10, the best-fitting model according to the fit measures that take model parsimony into account, imposes partial metric invariance (seven substantive loadings are freely estimated in various countries).

different constructs. We will start with a baseline model containing no method effects (M1), then consider several models with either a measured NARS factor (M2 M3) or inferred response scale factors (M4 M7), and finally present a model that incorporates both the best specification for the NARS factor and the best specification for the inferred response scale factors (M8). Third, for each type of method effect specification, we determine the best-fitting model (e.g., in terms of whether the method effects or method loadings are uniform across items, whether the inferred method factors corresponding to different response scales are correlated).

We start from a model in which locus, stability, and control are freely correlating latent constructs (traits) with three reflective indicators each. Country serves as a grouping variable. Several common model fit indices are reported in Table 5, and we primarily rely on the fit measures that trade off fit with parsimony (RMSEA, TLI, and BIC) when comparing models. In Models M1–M8, the trait factor loadings are freely estimated and no metric invariance is imposed (that is, the loadings are allowed to differ across countries). In the initial model, M1, no effects of NARS on the items are modeled and no method factors are specified. In Models M2 and M3, we evaluate NARS effects (without method factors); in Models M4–M7, we evaluate method factors (without NARS effects); and in Model M8 there are both NARS effects and method factors. Model M2 has three traits and unequal effects of NARS on the items. That is, for each country, the effect of NARS on each of the items is freely estimated. In Model M3a, NARS is specified to have the same effect on each of the nine items within each country. Closer inspection of the parameter estimates shows that the NARS effect on the second locus item (locus-b) was weaker than the effect on the other items. This result makes sense: Even though this item seems to use a scale format that is similar to the second item of the stability and control items with a scale range from 0 to 100, it is different because it uses a constant sum approach (see Table 4), which is inherently less prone to acquiescence (since respondents need to divide points over multiple options, making it impossible to be overly agreeable overall). Given these observations, we constrained the effect of NARS to be equal for only eight of the nine items in Model M3b (the exception is locus-b).

The next four models do not account for NARS effects but add (inferred) method factors. In particular, Model M4 has three traits and two correlated method factors but no effects of NARS on the items. As discussed, note that even though the second item of each trait uses a scale ranging from 0 to 100, the second locus item is fundamentally different from the other b-items, as it uses a constant sum format. Because of this, it would not be meaningful to estimate a method b factor, so we only included two method factors, one for method a and another for method c. Our approach was confirmed by the fact that a model with method effects for method b did not yield a meaningful method factor because the method loadings consisted of a mix of positive and negative loadings. Model M5 evaluates whether the correlation between the two method factors can be set to zero. Model M6 again has two correlated



**Figure 3.** Measurement model for the empirical application.

method factors but imposes the assumption of equal method loadings within each method factor. Model 7 combines both constraints in that it has a zero correlation between the method factors as well as equal loadings within each method factor. The model fit indices indicate that setting the method factor correlation to zero is not a plausible constraint, whereas constraining the loadings within each method factor to equality is justified. Model M8 simultaneously models method effects due to NARS and two inferred method factors using the preferred specification within each type of model: equal effects of NARS on eight of the items, with a separate effect for locus-b, and two correlated method factors with equal method loadings within each method factor (see Figure 3). As might be expected considering the preceding model comparisons, Model M8 has the best fit among all the models compared so far (with a higher CFI and TLI and lower RMSEA, SRMR, and BIC). Furthermore, although the chi-square test is significant, all alternative fit indices indicate acceptable, or nearly acceptable, fit (i.e., CFI and TLI > .95; RMSEA < .06, SRMR < .08).

Once the most appropriate method effects specification has been determined, one can test for metric invariance (i.e., in the current example, the focus is on the covariances between the factors, so scalar invariance is not required). Model M9 imposes full metric invariance on the trait factor loadings of Model M8 across countries. Because the decrease in model fit is modest, one could accept metric invariance, especially in light of the improvement in Bayesian information criterion (Steenkamp and Baumgartner 1998; Weijters, Puntoni, and Baumgartner 2017). However, if a researcher wants to be more conservative, one approach for arriving at a plausible model specification with partial metric invariance would be as follows. In a full metric invariance model, three loadings are set to 1 and six loadings are freely estimated but held invariant across all countries. In total,  $6 \times 13$  loadings could be freed. Using an alpha of .01 and a Bonferroni correction, the adjusted alpha would be .0001282. This would imply a critical value of

14.8 (rather than 3.84) for each individual modification index, and one could sequentially free noninvariant loadings until no modification index exceeds 14.8. We followed this procedure to arrive at Model M10, which resulted in the best-fitting model according to the fit measures that take model parsimony into account.

Table 6 reports a measurement analysis of the three attribution factors based on Model M1 (a naive model that does not account for NARS effects or method factors) versus Model M10 (which exhibits partial metric invariance and accounts for NARS effects and two method factors). For each construct, averages across the 14 countries in estimates of composite reliability and average variance extracted, as well as shared variance for each construct pair, are reported (Baumgartner and Weijters 2017). Note that for Model M10, composite reliability and average variance extracted refer to the internal consistency and average individual-item reliability of the substantive portion of the variance of each item only. The results show a consistent pattern: The composite reliability and average variance extracted values are slightly lower in the model in which method effects are accounted for (M10) than in the baseline model (M1), suggesting that the presence of method variance leads to somewhat inflated estimates of internal consistency if left uncontrolled (because shared method variance is incorrectly treated as substantive variance). The shared variance values are also lower in Model M10 than in the baseline model (M1), suggesting that the presence of method variance leads to inflated inter-factor correlations, which on the one hand could threaten discriminant validity (although this is not an issue in the current data) but on the other hand could also lead to exaggerated substantive relationships between constructs. The differences in estimates between the corrected and uncorrected models are not large in the present case, but if a researcher concluded that method effects were absent (using faulty methods) and proceeded to investigate the substantive hypotheses without controlling for CMV, the empirical findings would be ambiguous at best and misleading at worst.

In summary, the findings show that method variance is clearly present in the data. Both scale usage differences (or NARS) and commonalities for two of the response scales used for measuring the three attribution dimensions (methods a and c) contribute to shared method variance among the observed indicators. However, in the present case, common method variance does not change the substantive conclusions (i.e., the three attribution dimensions are significantly positively correlated regardless of whether or not method effects are accounted for) and the shared variance (squared correlation) between constructs is only reduced by 14% on average (across the 14 countries and three constructs) when method effects are taken into account.

## General Discussion

Common method variance is a concern in studies in which (1) the data come from a single source (often each respondent provides the data for all constructs measured in a study); (2)

**Table 6.** Composite Reliability, Average Variance Extracted, and Shared Variance for Models M1 and M10.

	Baseline Model Without Method Effects (Model M1)				Final Partial Metric Invariance Model with Method Effects (Model M10)			
	Locus	Stability	Control	Composite reliability	Locus	Stability	Control	Composite reliability
Locus	.612 [.591, .633]			.827 [.811, .843]	.578 [.551, .604]			.802 [.783, .821]
Stability	.278 [.196, .360]	.722 [.699, .745]		.898 [.885, .911]	.225 [.148, .302]	.696 [.666, .726]		.879 [.864, .895]
Control	.414 [.328, .501]	.358 [.260, .456]	.795 [.771, .818]	.929 [.922, .937]	.367 [.281, .453]	.317 [.227, .406]	.773 [.749, .797]	.911 [.901, .920]

Note: The table entries are averages across the 14 countries, with 95% confidence intervals shown in brackets. Average variances extracted are displayed on the diagonal, and shared variances between constructs (squared correlations) are displayed in the sub-diagonals. See Table 5 for details about Models M1 and M10.

the responses are based on the same response scale for all items (researchers often use a five- or seven-point Likert scale), none of which are reverse-keyed; and/or (3) the data are collected in the same setting at one point in time, with little separation between the dependent and independent variables. Under these circumstances, it is likely that the covariation between the measured items reflects not only the hypothesized substantive relationships between the constructs but also shared method variance. If this common method variance is not controlled, shared method variance may be mistaken for shared substantive variance.

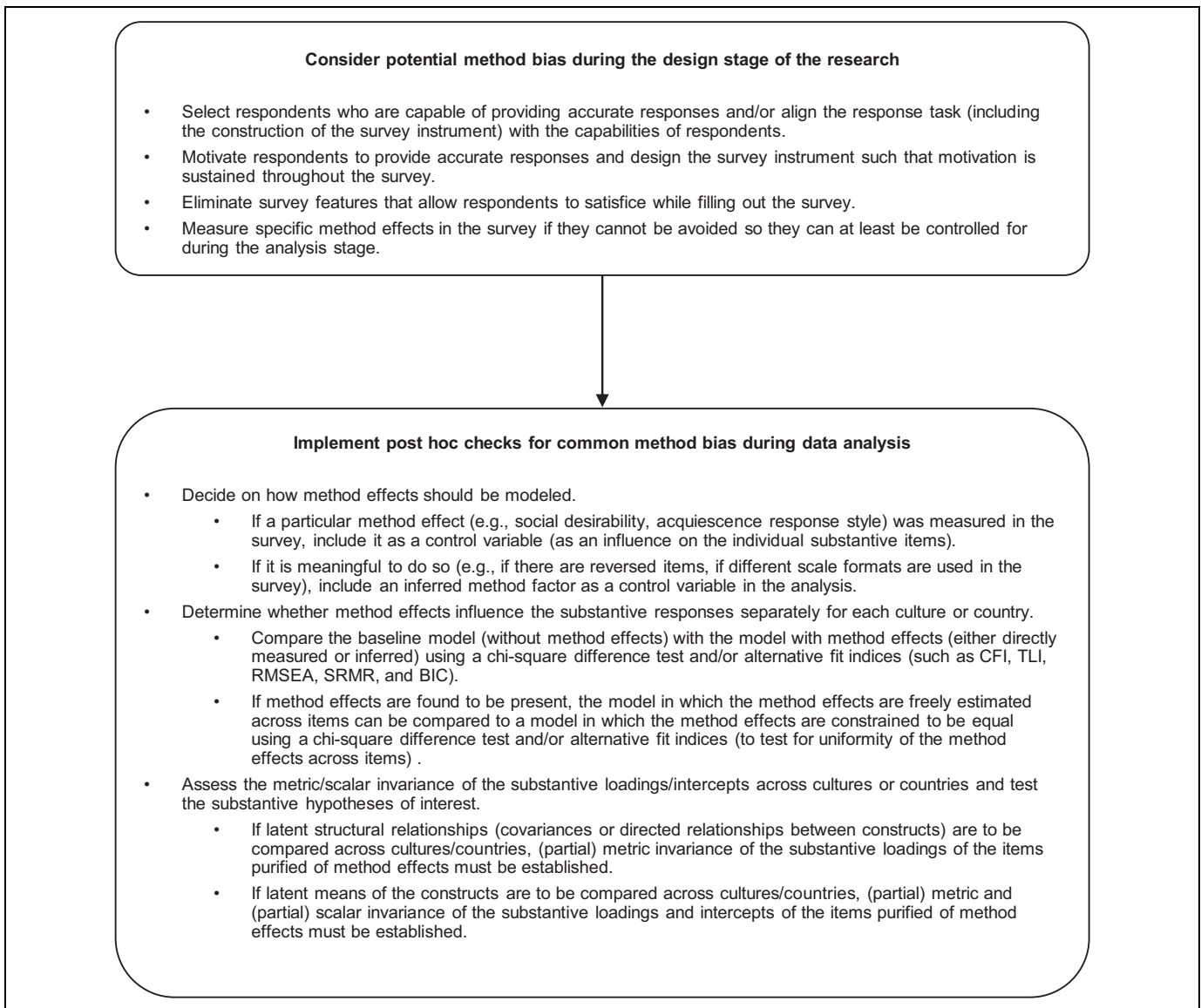
The conditions that give rise to common method variance point to strategies that researchers can adopt during the planning stages of research to avoid method effects. Because MacKenzie and Podsakoff (2012) already present a list of procedural remedies to counter method effects, we only provide a brief overview here (see also Podsakoff et al. 2003). MacKenzie and Podsakoff (2012) suggest that there are three sets of factors that increase the likelihood of method bias: lack of respondent ability to respond accurately, lack of respondent motivation to respond accurately, and item characteristics that encourage satisficing in surveys (rather than optimizing). Given these factors, distinct strategies for inhibiting method bias can be suggested. First, researchers should select respondents who are capable of providing the desired responses with a reasonable amount of effort and/or align the response task with the cognitive capabilities of respondents. With regard to the latter, researchers should implement the usual methods for improving the instructions, the formulation of the items, and the response scales (for details, see Baumgartner and Weijters 2019, Chapter 3). Second, researchers should take steps to increase respondents' motivation to answer all questions accurately (e.g., by emphasizing the relevance of the topics covered in the survey) and reduce survey characteristics that are detrimental to motivation (e.g., lengthy batteries of questions). One way to achieve the latter is to collect the required responses from multiple sources or in several stages. Third, researchers should eliminate factors that make it easy for respondents to satisfice, such as grouping related (especially very similar) items together (in which case respondents can provide the same answers without thinking) or using the same response scale for all items (which encourages respondents to choose the same scale position repeatedly). If the researcher believes that a survey is susceptible to specific common method effects and these effects cannot be avoided, then explicit measures of the method effects in question should be built into the survey so that they can be controlled for during the analysis stage.

Although procedural remedies at the design stage are very useful, they cannot guarantee that the data will be free of method bias. It is therefore necessary to check for method effects post hoc. Our review of *Journal of International Marketing* studies shows that international marketing researchers are cognizant of the threat of common method bias to the validity of their research findings, but it appears that they have often used suboptimal statistical methods for dealing with this

problem. Despite strongly worded recommendations that the Harman one-factor test not be used to ascertain the presence of CMV, it is apparently routinely reported in recently published articles in one of the premier international journals. The criticism leveled at the marker variable technique has been less harsh, but as shown in this article, the very concept of a marker variable is of doubtful validity, and the marker variables used in empirical studies cannot provide a meaningful control for CMV. Furthermore, the (usually untestable) assumptions underlying the technique are so unrealistic that the procedure lacks face validity. A moment's reflection should make it plain that marker variables should not be used to control for CMV: If a variable is conceptually unrelated to at least one of the substantive variables (and usually most of them) and does not capture method effects in any meaningful way (e.g., why should a demographic or other factual question be influenced by method effects?), how could such a variable be useful in accounting for method effects? And if the smallest (or second-smallest) correlation is close to zero, as it often is, it is a foregone conclusion that partialing out this correlation will not affect the substantive correlations. The only reason for the popularity of this technique seems to be that it usually leads to the desired conclusion that no method variance is present in the data. The same is also true for the Harman one-factor test.

Although we believe that researchers often discount the threat of CMV because of faulty procedures, we do not want to argue that CMV will invariably invalidate research findings. Even if CMV is present in empirical data (which need not be the case), it does not necessarily change the substantive conclusions. However, a researcher should (1) test whether observed measures contain a significant amount of method variance and (2) ascertain whether taking method effects into account significantly alters substantive relationships of interest. To this end, we propose a post hoc procedure for incorporating method effects in international marketing research. The procedure uses either a directly measured method effect or an inferred method factor as a control variable at the item level. If a latent method factor is specified, a researcher must make sure that the presumed method factor actually captures method effects, which usually requires a research design that makes the specification of a method factor meaningful (e.g., balanced scales with an equal number of regular and reversed items). The proposed procedure makes it possible to investigate method effects across cultures or countries in a principled way and uses multisample structural equation modeling and invariance testing.

Figure 4 presents a flow chart of how researchers can deal with CMV in international marketing research. It consists of both an a priori stage during the design of a research study (which we did not emphasize in this article) and a post hoc stage during the analysis of the data collected in the study, as discussed in this article and exemplified in our empirical study. It is our hope that the proposed procedure will prove useful in further studies and that, in the future, international marketing researchers will employ this and similar sophisticated approaches to deal with the problem of CMV.



**Figure 4.** Proposed procedure for dealing with method effects in international marketing research.

## Appendix A

We used the following CRM scenarios in the empirical study:

- During the month of November, people from different European countries will collect money from consumers and companies. Among others, they will call on [the brand] to support their initiative.
- During the month of November, all cans of [the brand] will carry a reference to an organization. Via the stated bank account number, consumers can transfer a contribution.
- During the month of November, employees of [the brand] will get to work as volunteers in their time off.
- During the month of November, employees of [the brand] will get to work as volunteers in their time off. Moreover, consumers can give sponsoring for every hour of volunteered time that is provided that way.
- [The brand] will give away part of its profit in November.

- In November, [the brand] will give away a fixed amount of money for every can that consumers buy.
- In the month of November, employees of [the brand] will get to work as volunteers during their (paid) working hours.
- In the month of November, employees of [the brand] will get to work as volunteers during their (paid) working hours. Moreover, consumers can give sponsoring for every hour of volunteered time that is provided that way.

## Acknowledgments

The authors would like to thank the reviewers for many helpful comments.

## 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 first author gratefully acknowledges support from the Smeal Chair endowment.

## References

- Baumgartner, Hans and Jan-Benedict E.M. Steenkamp (2001), "Response Styles in Marketing Research: A Cross-National Investigation," *Journal of Marketing Research*, 38 (2), 143–56.
- Baumgartner, Hans and Bert Weijters (2015), "Response Biases in Cross-Cultural Measurement," in *Handbook of Culture and Consumer Behavior*, Sharon Ng and Angele Y. Lee, eds. Oxford, UK: Oxford University Press, 150–80.
- Baumgartner, Hans and Bert Weijters (2017), "Structural Equation Modeling," in *Advanced Methods for Modeling Markets*, Peter S. H. Leeflang, Jaap E. Wieringa, Tammo H. A. Bijmolt and Koen H. Pauwels, eds. Cham, Switzerland: Springer International Publishing AG, 335–60.
- Baumgartner, Hans and Bert Weijters (2019), "Measurement in Marketing," *Foundations and Trends® in Marketing*, 12 (4), 278–400.
- Baumgartner, Hans, Bert Weijters, and Rik Pieters (2021), "The Biasing Effect of Common Method Variance: Some Clarifications," *Journal of the Academy of Marketing Science*, 49 (2), 221–35.
- Ellen, Pam Scholder, Deborah J. Webb, and Lois A. Mohr (2006), "Building Corporate Associations: Consumer Attributions for Corporate Socially Responsible Programs," *Journal of the Academy of Marketing Science*, 34 (2), 147–57.
- Greenleaf, Eric A. (1992), "Measuring Extreme Response Style," *Public Opinion Quarterly*, 56 (3), 328–50.
- Hulland, John, Hans Baumgartner, and Keith Marion Smith (2018), "Marketing Survey Research Best Practices: Evidence and Recommendations from a Review of JAMS Articles," *Journal of the Academy of Marketing Science*, 46 (1), 92–108.
- Klein, Jill and Niraj Dawar (2004), "Corporate Social Responsibility and Consumers' Attributions and Brand Evaluations in a Product-Harm Crisis," *International Journal of Research in Marketing*, 21 (3), 203–17.
- Lindell, Michael K. and David J. Whitney (2001), "Accounting for Common Method Variance in Cross-Sectional Research Designs," *Journal of Applied Psychology*, 86 (1), 114–21.
- MacKenzie, Scott B. and Philip M. Podsakoff (2012), "Common Method Bias in Marketing: Causes, Mechanisms, and Procedural Remedies," *Journal of Retailing*, 88 (4), 542–55.
- Marsh, Herbert W. (1989), "Confirmatory Factor Analyses of Multitrait-Multimethod Data: Many Problems and a Few Solutions," *Applied Psychological Measurement*, 13 (4), 335–61.
- Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff (2003), "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, 88 (5), 879–903.
- Podsakoff, Philip M., Scott B. MacKenzie, and Nathan P. Podsakoff (2012), "Sources of Method Bias in Social Science Research and Recommendations on How to Control It," *Annual Review of Psychology*, 63, 539–69.
- Richardson, Hettie A., Marcia J. Simmering, and Michael C. Sturman (2009), "A Tale of Three Perspectives: Examining Post Hoc Statistical Techniques for Detection and Correction of Common Method Variance," *Organizational Research Methods*, 12 (4), 762–800.
- Sen, Sankar and Chitra Bhanu Bhattacharya (2001), "Does Doing Good Always Lead to Doing Better? Consumer Reactions to Corporate Social Responsibility," *Journal of Marketing Research*, 38 (2), 225–43.
- Steenkamp, Jan-Benedict E. M. and Hans Baumgartner (1998), "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research*, 25 (1), 78–90.
- Steenkamp, Jan-Benedict E. M., Martijn G. De Jong, and Hans Baumgartner (2010), "Socially Desirable Response Tendencies in Survey Research," *Journal of Marketing Research*, 47 (2), 199–214.
- Tellis, Gerard J. and Deepa Chandrasekaran (2010), "Extent and Impact of Response Biases in Cross-National Survey Research," *International Journal of Research in Marketing*, 27 (4), 329–41.
- Van Auken, Stuart, Thomas E. Barry, and Richard P. Bagozzi (2006), "A Cross-Country Construct Validation of Cognitive Age," *Journal of the Academy of Marketing Science*, 34 (3), 439–55.
- Varadarajan, P. Rajan and Anil Menon (1988), "Cause-Related Marketing: A Coalignment of Marketing Strategy and Corporate Philanthropy," *Journal of Marketing*, 52 (3), 58–74.
- Viswanathan, Madhu (2005), *Measurement Error and Research Design*. Thousand Oaks, CA: SAGE Publishing.
- Weijters, Bert, Hans Baumgartner, and Maggie Geuens (2016), "The Calibrated Sigma Method: An Efficient Remedy for Between-Group Differences in Response Category Use on Likert Scales," *International Journal of Research in Marketing*, 33 (4), 944–60.
- Weijters, Bert, Hans Baumgartner, and Niels Schillewaert (2013), "Reversed Item Bias: An Integrative Model," *Psychological Methods*, 18 (3), 320–34.
- Weijters, Bert, Niels Schillewaert, and Maggie Geuens (2008), "Assessing Response Styles Across Modes of Data Collection," *Journal of the Academy of Marketing Science*, 36 (3), 409–22.
- Weijters, Bert, Stefano Puntoni, and Hans Baumgartner (2017), "Methodological Issues in Cross-Linguistic and Multilingual Advertising Research," *Journal of Advertising*, 46 (1), 115–28.
- Weiner, Bernard (1980), "A Cognitive (Attribution)-Emotion-Action Model of Motivated Behavior: An Analysis of Judgments of Help-Giving," *Journal of Personality and Social Psychology*, 39 (2), 186–200.
- Weiner, Bernard (1985), "An Attributional Theory of Achievement Motivation and Emotion," *Psychological Review*, 92 (4), 548–73.
- Weiner, Bernard, Sandra Graham, and Carla Chandler (1982), "Pity, Anger, and Guilt: An Attributional Analysis," *Personality and Social Psychology Bulletin*, 8 (2), 226–32.
- Williams, Larry J., Nathan Hartman, and Flavia Cavazotte (2010), "Method Variance and Marker Variables: A Review and Comprehensive CFA Marker Technique," *Organizational Research Methods*, 13 (3), 477–514.
- Wong, Nancy, Aric Rindfleisch, and James E. Burroughs (2003), "Do Reverse-Worded Items Confound Measures in Cross-Cultural Consumer Research? The Case of the Material Values Scale," *Journal of Consumer Research*, 30 (1), 72–91.