ANNUAL REVIEWS

Annual Review of Organizational Psychology and Organizational Behavior

Common Method Bias: It's Bad, It's Complex, It's Widespread, and It's Not Easy to Fix

Philip M. Podsakoff,[1] Nathan P. Podsakoff,[2] Larry J. Williams,[3] Chengquan Huang,[1] and Junhui Yang[1]

[1]Warrington College of Business, University of Florida, Gainesville, Florida, USA; email: philip.podsakoff@warrington.ufl.edu

[2]Eller College of Management, University of Arizona, Tucson, Arizona, USA

[3]Jerry S. Rawls College of Business, Texas Tech University, Lubbock, Texas, USA

Annu. Rev. Organ. Psychol. Organ. Behav. 2024. 11:17–61

First published as a Review in Advance on November 29, 2023

The *Annual Review of Organizational Psychology and Organizational Behavior* is online at orgpsych.annualreviews.org

https://doi.org/10.1146/annurev-orgpsych-110721-040030

Keywords

common method variance, same-source bias, Harman's single-factor test, marker variable technique, procedural and statistical remedies

Abstract

Despite recognition of the harmful effects of common method bias (CMB), its causes, consequences, and remedies are still not well understood. Therefore, the purpose of this article is to review our current knowledge of CMB and provide recommendations on how to control it. We organize our review into five main sections. First, we explain the harmful effects of CMB (why it is bad). Second, we discuss the complexity caused by the fact that there are multiple sources of CMB, several of which are likely to be present in any study. Third, we present evidence that the conditions under which CMB is likely to occur are relatively widespread, and fourth, we explain why CMB is not easy to fix. Finally, we identify several avenues for future research.

> Any measure. . .reflects not only a theoretical concept of interest but also measurement error. Measurement error. . .can be partitioned into random error and systematic error, such as method variance. Method variance refers to variance attributable to the measurement method rather than the construct of interest. . . . Each of these two components can have serious confounding influences on empirical research and yield misleading conclusions. . . . Because measurement errors (i.e., random error and method variance) provide potential threats to the validity of research findings, it is important to validate measures and disentangle the distorting influences of these errors before testing theory.
>
> —Bagozzi et al. (1991, p. 421)

INTRODUCTION

As indicated by the quotation above, researchers in the organizational and behavioral sciences are concerned about the potential harmful effects of systematic measurement error on the validity of their research findings. Indeed, since Campbell & Fiske (1959) focused attention on this issue 65 years ago, it has been an enduring topic of discussion (e.g., Cote & Buckley 1987, Doty & Glick 1998, Evans 1985, Hulland et al. 2018, Podsakoff et al. 2003, Spector et al. 2019). However, despite widespread recognition that method biases can have several harmful effects, our reading of the literature and discussions with colleagues suggest that the causes and consequences of common method bias (CMB), and the remedies for dealing with it, are still not well understood, for several reasons. First, the effects that method factors have on the observed relationships between variables are complex. Indeed, several sources of CMB exist, and it is difficult if not impossible to control all of them in a single study. This complexity is compounded by the fact that as the field has moved from examining simple research designs using regression techniques to more complicated multidimensional constructs (Johnson et al. 2011) using multilevel analyses (Lai et al. 2013, Mathieu et al. 2012), it has become more difficult to determine how to control for the effects of CMB.

Second, there is considerable debate about the potential impact that CMB has on the relationships between constructs. Some scholars argue that CMB is an important problem that needs to be identified and controlled (e.g., Bagozzi 2011; Burton-Jones 2009; Cote & Buckley 1988; Podsakoff et al. 2003, 2012; Williams et al. 2010), and others claim that the potential effects of CMB are (at best) exaggerated (Bozionelos & Simmering 2022, Fuller et al. 2016, Spector 1987, Spector & Brannick 2010) and (at worst) may represent an urban legend that can generally be ignored (Brannick et al. 2010, question 2; Spector 2006). This debate produces confusion about the consequences of CMB and the necessity of identifying remedies to control them.

Third, the sheer amount of material published on CMB makes it difficult for even the most devoted scholars to keep up with this literature. For example, more than three dozen articles examining techniques for assessing or controlling CMB have been published since 2010 (e.g., Antonakis et al. 2010, Jordan & Troth 2020, Kock et al. 2021, MacKenzie & Podsakoff 2012, Yao & Xu 2021, Zhang et al. 2022), and these articles received more than 13,000 combined citations in 2022 alone (according to a 2023 Google Scholar search). Finally, even when researchers are aware of the potential problems that systematic CMB can produce, they may be unclear about how to minimize their harmful effects.

Therefore, the goal of this article is to increase our understanding of CMB's causes, consequences, and potential remedies. First, we discuss why CMB is bad, highlighting the harmful effects it can have on the estimates of construct reliability and validity and on the relationships between measures of different constructs. Second, we explore the complex nature of CMB. To better understand this complexity and what it means for researchers interested in controlling CMB, we discuss its sources and the conditions under which it is likely to have its biggest effects. Third, we summarize evidence indicating that studies using designs susceptible to CMB are relatively widespread. Indeed, the evidence suggests that the conditions in which CMB is likely to have biasing effects are quite common in several disciplines. Fourth, we discuss why CMB is not easy to fix by reviewing the strengths and limitations of various procedural and statistical remedies used to minimize its harmful effects. Finally, we identify several avenues for future research.

This article builds on and extends our earlier reviews (MacKenzie & Podsakoff 2012; Podsakoff et al. 2003, 2012; Podsakoff & Organ 1986) and makes several contributions to the literature. First, we provide an updated review incorporating research on CMB reported since our earlier article (Podsakoff et al. 2012). Second, we provide illustrations to help clarify the potential effects that CMB has on the reliability and validity of measures, as well as the relationships between constructs. Third, we examine and critique some recent claims (Bozionelos & Simmering 2022, Cruz 2022, Fuller et al. 2016) that CMB is not a threat to research findings. Finally, we provide recommendations to guide researchers interested in controlling the potential effects of CMB and discuss several avenues for future research.

COMMON METHOD BIAS IS BAD

There are two harmful effects of method variance (Bagozzi 2011; Baumgartner & Steenkamp 2001; Cote & Buckley 1987, 1988; Doty & Glick 1998; MacKenzie & Podsakoff 2012; Podsakoff et al. 2003, 2012; Williams et al. 2010). The first is that method variance can bias estimates of the reliability and validity of a latent variable. The second is that uncontrolled method variance can bias parameter estimates of the relationships between measures of different constructs. To develop a better understanding of these biases, let us examine the measurement model in **Figure 1**, which depicts a latent construct and the reflective indicators (items) used to measure it. There are several points worth noting about **Figure 1**. First, the latent construct accounts for only a portion of the variance in the indicators, and the remaining variance in these items is accounted for by systematic and random error. Second, the percentage of variance accounted for by the latent construct varies across indicators. Although the variance in some indicators is accounted for primarily by the underlying construct, this is not true for all the indicators. This reflects the

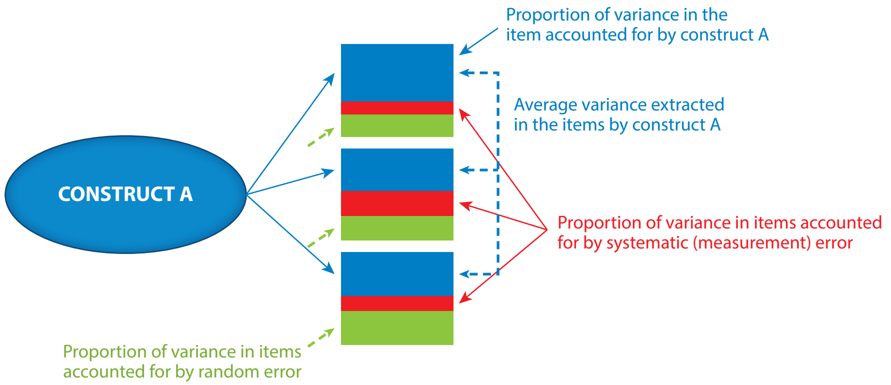

Figure 1

Partitioning indicator variance into component parts. Consistent with conventions in the literature (e.g., Bollen 1989, Brown 2015), the oval represents a latent construct and the rectangles represent reflective indicators (items) used to measure the construct. The different colors represent the different sources of variance accounted for in each indicator. Specifically, the variance in each indicator is partitioned into portions attributable to the construct being measured (*blue*), the method used to measure the construct (*red*), and random measurement error (*green*). The dotted blue lines reflect the average variance extracted in the items by the underlying construct. For simplicity, we omit variance that may be unique to the items.

notion that some indicators are closer representations of the core meaning of a latent construct than others, and that the variance accounted for in these items by the construct is greater than the variance accounted for in indicators that are less central to the construct's meaning (Little et al. 1999, MacKenzie et al. 2011). Third, the average variance extracted (AVE) in the items by the underlying construct is used as one indication of the quality of the measures. More specifically, Fornell & Larcker (1981) argue that it is desirable for the AVE of a latent variable to be greater than 0.50, because that indicates that most of the variance in the measures is accounted for by the underlying latent variable, rather than by measurement error.

Note also that the proportion of variance attributable to method variance varies across the items. This observation is consistent with the findings of several studies, which have reported that method factors have unequal effects on different measures, regardless of whether they are different measures of the same construct (Johnson et al. 2011; Podsakoff et al. 2012; Rafferty & Griffin 2004, 2006; Williams et al. 2010) or measures of different constructs (Baumgartner & Steenkamp 2001, Cote & Buckley 1987). Indeed, indicators can be influenced by different method factors, and the same method factor can have stronger or weaker effects on different indicators of a given construct or on different constructs (Campbell & Fiske 1959, McDermott & Sharma 2017, Messick 1991, Podsakoff et al. 2012, Spector et al. 2022). Finally, **Figure 1** shows that random measurement error also varies across the items used to measure the focal construct. This variation reflects the fact that this form of error is random and depends on the amount of variance accounted for by the latent construct and the methods used to measure it.

Effects of Method Factors on Estimates of the Reliability and Validity of Latent Constructs

Figure 2 illustrates the harmful effects that systematic CMB can have on a latent construct's reliability and validity. Here, three indicators of the focal (latent) construct are measured from the same source, and as a result, systematic variance that is attributable to the method of measurement is lumped in with systematic variance that is attributable to the latent construct itself. The figure captures systematic variance in the items that is attributable to both the latent construct and

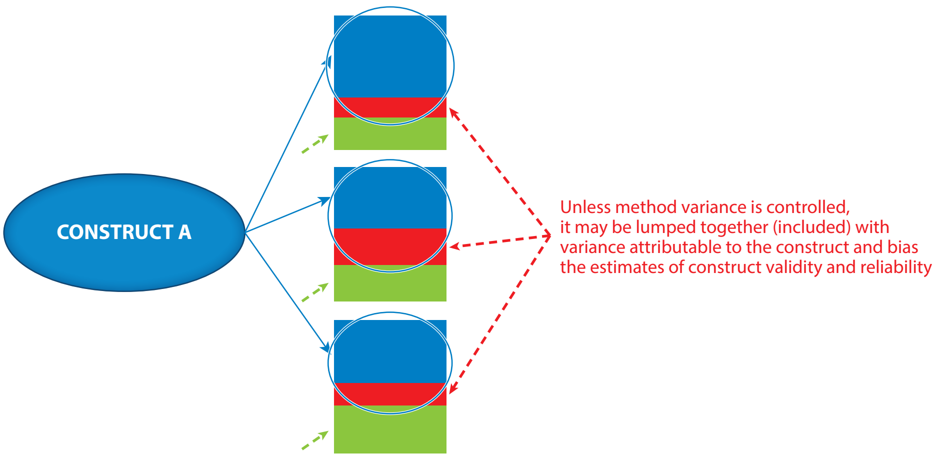

Figure 2

Example of systematic measurement error biasing estimates of construct validity and reliability. The blue circles represent systematic variance in the items that is attributable to the latent construct and any method characteristics shared by the items.

any method characteristics shared by the items. When variance attributable to method factors is not controlled and is lumped together with construct variance, it can bias the estimates of construct validity and reliability. The biasing effects of systematic measurement error due to common methods on estimates of construct validity and reliability produce several potential problems.

First, method factors that inflate or attenuate interitem covariation will bias estimates of factor loadings, reliability coefficients, and AVE estimates (MacKenzie & Podsakoff 2012, Podsakoff et al. 2012), which can lead to incorrect conclusions about the adequacy of a scale's reliability and item-level convergent validity (Bagozzi 1984; Baumgartner & Steenkamp 2001; Brannick et al. 2010; Cote & Buckley 1987; Podsakoff et al. 2003, 2012; Williams et al. 1989, 2010). Indeed, when systematic method variance in the items is not controlled, researchers may overestimate the reliability and validity of their scales and, in more extreme cases, conclude that their scales are reliable and valid when in fact they are not. Second, common method variance (CMV) can produce inaccurate "corrected" correlations in meta-analyses when the reliability estimates used to calculate the correction are biased (Le et al. 2009, MacKenzie & Podsakoff 2012, Podsakoff et al. 2012). Specifically, reliability-corrected correlations will understate the relationship between focal variables when reliability estimates are inflated by CMB, and these estimates will overstate the relationship between these variables when reliability estimates are attenuated by CMB.

Evidence of the Effects of Common Method Bias on Construct Validity and Reliability

Table 1 summarizes the evidence of the biasing effect of method factors on construct validity and reliability. This table reports the results from seven meta-analytic studies that applied confirmatory factor analysis (CFA) techniques to previously published multitrait multimethod (MTMM) matrices to estimate the proportion of trait, method, and error variance that is present in the data. As shown in the table, almost a quarter (24%) of the total variance in the items used in these studies is due to method factors. When compared with the recommended 0.50 cutoff value for AVE (Fornell & Larcker 1981), these estimates speak to the biasing potential of method factors on reliability and validity estimates. For example, if the AVE for a latent construct is 0.65, and 24% of the total variance in the items (on average) is attributable to method factors that are not controlled for, then the actual amount of variance in the items that is attributable to the latent construct is only 41% ($0.65 - 0.24 = 0.41$). This example illustrates how CMB could cause researchers to overestimate the validity and reliability of their measures, even when the actual AVE is substantially below the recommended cutoff value.

Method Factors Can Bias Estimates of the Relationships Between Constructs

As illustrated in **Figure 3**, uncontrolled method factors can also have harmful effects on the parameter estimates representing the empirical relationships between measures of two or more constructs. This example shows two correlated latent variables (supportive leader behavior and employee helping behavior), each measured by three indicators using self-reports obtained from the same source (employees). Though the researcher is interested in examining the "true" relationship between these latent constructs, to the extent that their measures share method variance that is not controlled, the observed correlation between them will be biased.

Although the literature typically emphasizes inflation of parameter estimates, several researchers (Baumgartner et al. 2021; Cote & Buckley 1988; MacKenzie & Podsakoff 2012; Podsakoff et al. 2003, 2012; Siemsen et al. 2010) have noted that CMV can also deflate or have no net effect on estimates of the relationship between two constructs. Specifically, Cote & Buckley (1988) note that method factors (*a*) deflate the relationship between the measures

Table 1 Summary of studies reporting CFAs of MTMM matrices to partition trait, method, and error variance in latent variables

Reference	Sample	Estimation technique	Variance attributable to trait (construct) factors	Variance attributable to method factors	Variance attributable to random error
Cote & Buckley (1987)	70 matrices examining a wide variety of constructs from the fields of marketing, psychology/sociology, education, and other business disciplines	Traditional CFA of MTMM matrices (CFA-MTMM)	42%	26%	32%
Williams et al. (1989)[a]	11 matrices involving perceptions of jobs and work environments	Traditional CFA of MTMM matrices (CFA-MTMM)	48%	25%	21%
Buckley et al. (1990)	61 matrices examining a variety of constructs	Traditional CFA of MTMM matrices (CFA-MTMM)	42%	22%	36%
Doty & Glick (1998)	28 matrices examining constructs from a variety of social science disciplines	Traditional CFA of MTMM matrices (CFA-MTMM)	46%	32%	22%
Mishra (2000)	6 matrices examining health care–related constructs	Traditional CFA of MTMM matrices (CFA-MTMM)	36%	30%	34%
Ketokivi & Schroeder (2004)	Data from 164 manufacturing plants from five countries (Germany, Italy, Japan, United Kingdom, United States) in three industries (automotive supply, machinery, electronics) on four measures of market performance (price, product quality, product image, product features) taken from three sources (plant manager, plant superintendent, plant research coordinator)	CFA of MTMM matrices using a CTUM	36%	18%[b]	46%
Lance et al. (2010)	18 matrices	CFA of MTMM matrices using a CTCM	40%	18%	42%
Averages			41%	24%	33%

[a]Values reported for variance estimates represent medians.
[b]The average amount of method variance per source varied depending on the performance measure used (e.g., average method variance for price = 4%; quality = 25%; product image = 20%; product features = 23%), providing additional evidence that the amount of method variance differs across constructs.
Table adapted with permission from Podsakoff et al. (2012).
Abbreviations: CFA, confirmatory factor analysis; CTCM, correlated traits–correlated methods model; CTUM, correlated traits–uncorrelated methods model; MTMM, multitrait multimethod.

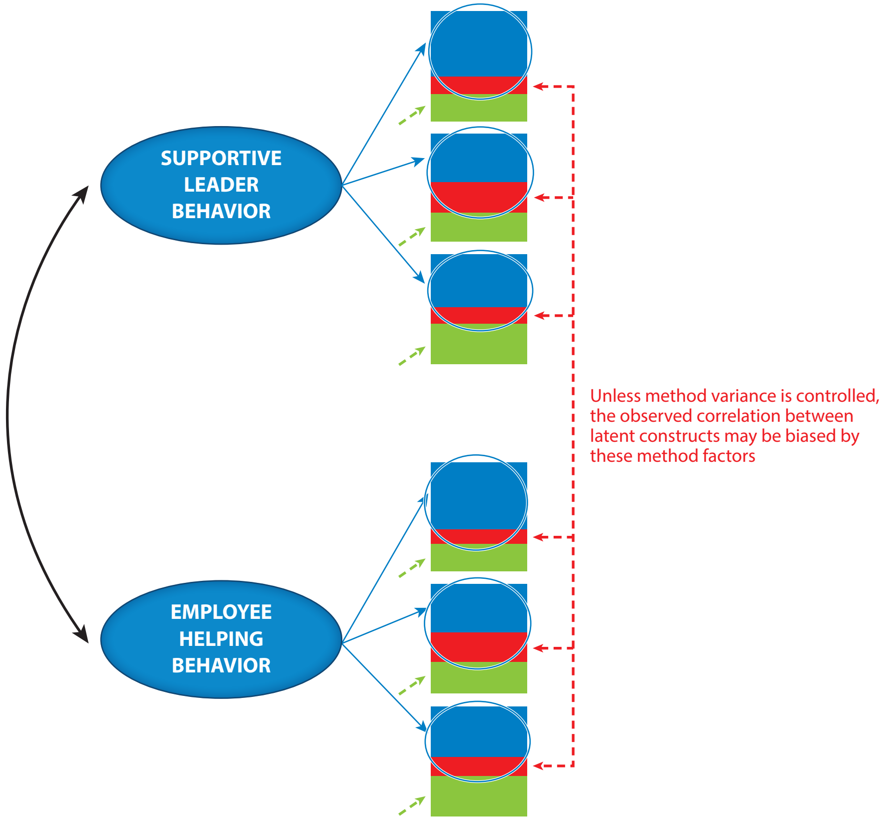

Figure 3

Example of systematic measurement error biasing the relationship between constructs. Each of the two correlated latent variables (supportive leader behavior and employee helping behavior) is measured by three indicators using self-reports obtained from the same source (employees).

of two constructs when the correlation between the method factors is lower than the observed correlation between the measures with method effects removed, (*b*) inflate the relationship when the correlation between the method factors is higher than the observed correlation between the measures with method effects removed, and (*c*) have no influence on the relationship when the correlation between the method factors is the same as the observed correlation between the variables with the method effects removed. However, regardless of whether the method factor inflates or deflates the relationship, it can cause several serious problems (Bagozzi 1984; Baumgartner & Steenkamp 2001; MacKenzie & Podsakoff 2012; Podsakoff et al. 2003, 2012; Siemsen et al. 2010).

First, since systematic measurement error can inflate or deflate the estimates of the observed relationships between two latent variables, these errors may lead a researcher to conclude either that a relationship between the two variables exists when it does not (Type I error) or that a relationship does not exist when indeed it does (Type II error). Second, if the predictor and criterion variables share systematic error variance, the amount of variance accounted for in the criterion variable(s) by the predictor variable(s) may be either understated or overstated. Finally, because systematic measurement error can inflate or deflate the estimates of the observed relationships between latent variables, it can enhance or attenuate the observed relationships between a focal construct and its antecedents, correlates, and consequences, and subsequently influence the inferences made about the construct's convergent, discriminant, nomological, and/or criterion-related validity.

Evidence of the Effects of Common Method Bias on the Covariation Between Constructs

A substantial amount of evidence indicates the potential effects that CMB has on the covariation between measures of constructs. Some of this evidence comes from research on the differences in the observed relationships reported when the measures of the predictor and criterion variables are obtained from the same (as opposed to different) raters, or the potential effects that research designs (cross-sectional versus lagged designs) have on these relationships. Still other evidence comes from research on the effects that item characteristics, item contexts, or measurement contexts have on CMB. As a starting point, we compare the differences in the correlations between variables when they were obtained from the same (versus different) sources and at the same (versus different) times (we return to the effects of other factors in the section titled The Effects of Common Method Bias Are Complex). To obtain estimates of the effects of CMB on the covariation between constructs, we analyzed correlations reported in published meta-analyses. We searched for and coded meta-analyses that reported correlations between measures of constructs rated by the same and different sources or rated at the same and different times. Our analyses included 233 bivariate correlations from 59 meta-analyses for the effects of rating sources as well as 236 bivariate correlations from 33 meta-analyses for the effects of rating times (for details on the literature search, inclusion criteria, and coding and the list of meta-analyses included in our analyses, see **Supplemental Appendix A**).

Supplemental Material >

Although several researchers (e.g., Baumgartner et al. 2021, Podsakoff et al. 2003) have observed heterogeneity in the effects of CMB, few studies have attempted to identify the factors that predict this variability. To help determine whether correlations of various relationships are subject to CMB to the same extent, we coded the valence (positive versus negative) of the predictor and criterion variables of each relationship in our data set and examined how it influenced the observed correlations. Earlier research (Kam & Meyer 2015, Magazine et al. 1996, Zeng et al. 2020) showed that positively and negatively worded items influence both construct dimensionality and nomological validity. Examples of predictor variables categorized as having a positive valence include positive individual differences [e.g., conscientiousness, agreeableness, core self-evaluation (CSE)], work attitudes (e.g., job satisfaction, organizational commitment), task characteristics (e.g., job challenge, autonomy), and leadership behaviors (e.g., transformational leadership, servant leadership). In contrast, examples of predictor variables categorized as having a negative valence include negative individual differences (e.g., neuroticism, negative affect), work attitudes (e.g., burnout, cynicism about change), leadership behaviors (e.g., abusive supervision, unethical leadership), and job stressors or strains. Positive criterion variables include positive attitudes (e.g., commitment, satisfaction, engagement) or behaviors/performance [e.g., task performance, organizational citizenship behavior (OCB), creativity]. Negative criterion variables include negative attitudes (e.g., negative affect), perceptions (e.g., role ambiguity, role conflict), and behaviors [e.g., counterproductive work behavior (CWB)]. For a complete list of variables included in each valence category, see **Supplemental Appendix B**.

After coding the valence of the predictor and criterion variables, we examined four categories of relationship valence: positive–positive, positive–negative, negative–negative, and negative–positive. For each of these categories, we obtained the weighted average inflation rates in the correlations of same-source ratings compared with different-source ratings as well as in the correlations of cross-sectional designs compared with lagged designs (i.e., $\rho_{\text{same source}}/\rho_{\text{different source}}$

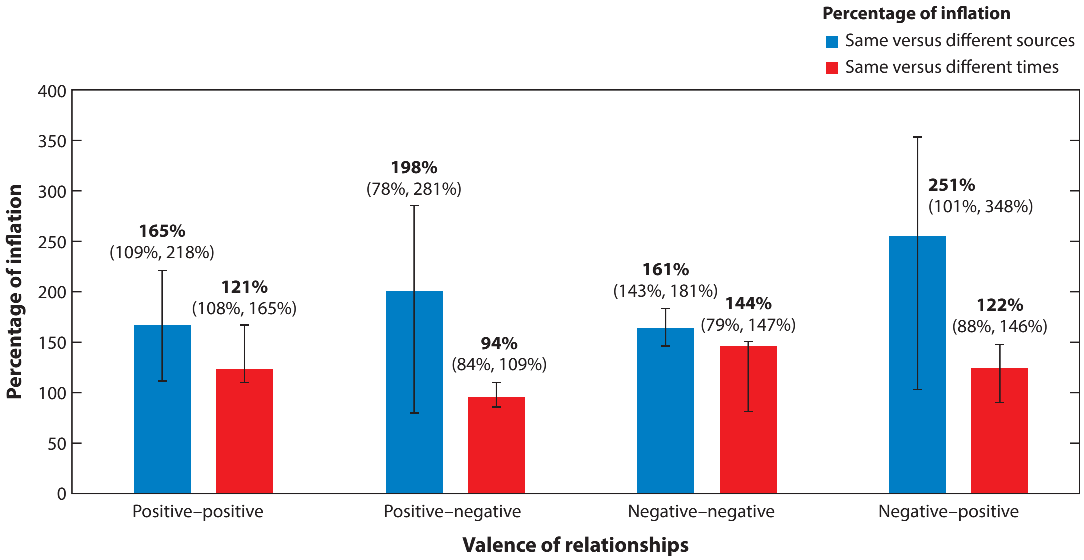

Figure 4

Effects of rating source and temporal separation on correlations across relationships of different valence. The number above each bar is the average inflation rate, followed by its twentieth-to-eightieth-percentile range. The blue bars compare the weighted average correlations of same-source versus different-source ratings across valence categories. The red bars compare the weighted average correlations of same-time versus different-time ratings across valence categories. For details on the meta-analytic data used in these estimates, see **Supplemental Appendix C**, **Table C1** (same versus different sources) and **Table C2** (same versus different times).

or $\rho_{\text{same time}}/\rho_{\text{different time}}$), weighted by the number of studies included in the meta-analyses we coded.

Effects of Obtaining Measures from the Same Versus Different Sources

Supplemental Material >

Podsakoff et al. (2012) have noted that obtaining predictor and criterion variables from different sources prevents the respondent's mindset or mood from influencing the ratings of these variables. Thus, ratings obtained from different sources should contain less CMB than ratings obtained from the same source. **Figure 4** compares the average weighted correlations of same-source versus different-source ratings across valence categories, showing the average inflation rate and the twentieth-to-eightieth-percentile range for each valence relationship pair. Several points regarding these data are worth noting. First, the average inflationary effects are above 160% for all relationship pairs, indicating that the correlations between the constructs included in these analyses are (on average) biased on the basis of rating source. Second, the inflationary effects of rating source vary across valence pair categories. Whereas the average inflation rate for negative–positive relationship pairs is 251% and the average inflation rate for positive–negative relationship pairs is 198%, the average inflation rates for positive–positive and negative–negative relationship pairs are 165% and 161%, respectively. This finding suggests that the average inflationary effects based on ratings source for relationship pairs that include both positive and negative valence constructs are somewhat greater than the average inflationary effects based on construct pairs that include only positive (negative) valence. Third, **Figure 4** indicates that there is substantial variability in the amount of inflationary effects across the pairs, again with more variability in the positive–negative and negative–positive construct pairs (with inflationary percentage ranges for the twentieth to eightieth percentile from 78% to 281% and 101% to 348%, respectively) than in the positive–positive or negative–negative construct pairs (with inflationary ranges for the twentieth to eightieth percentile from 109% to 218% and 143% to 181%, respectively).

Taken together, these findings suggest that (*a*) although rating source does (on average) influence the amount of the potential inflationary effects of CMB on predictor–criterion variable relationship pairs, (*b*) the valence of the construct pairs also influences the potential inflationary effects of CMB based on the rating source. Finally, the variability in the inflationary effects across these relationship pairs suggests that other factors (e.g., the specific variables included in the analyses, or sampling error) influence the size of these effects.

Effects of Obtaining Measures at the Same Versus Different Times

Researchers (N.P. Podsakoff et al. 2013; P.M. Podsakoff et al. 2003, 2012) have noted that temporal separation should decrease CMB because it reduces respondents' ability and motivation to access responses to previous items and to maintain consistency when responding to subsequent items. This observation suggests that correlations between variables obtained at the same point in time should be larger than correlations between the same variables taken at different points in time. **Figure 4** shows comparisons of the average weighted correlations of same-time versus different-time ratings across relationship categories. This figure indicates that although all but the positive–negative valence relationship pairs show an average inflationary effect when the predictor and criterion variables are obtained at the same (versus different) time periods, the average inflationary effects for positive–positive and negative–positive construct pairs are about the same (121% and 122%, respectively), and the average inflationary effect for the negative–negative relationship pairs is somewhat greater (144%). Taken together with the effects of rating source, these findings suggest that temporal separation of predictor and criterion variables generally does not have as big an effect on the estimates of these relationships as does obtaining ratings from different sources. Nevertheless, consistent with expectations that temporal separation should reduce respondents' opportunities to use previous responses to questions when responding to subsequent questions, the results reported in **Figure 4** indicate that such separation generally reduces the observed relationships between predictor and criterion variables.

THE EFFECTS OF COMMON METHOD BIAS ARE COMPLEX

Several other sources of CMB have been identified in the literature (MacKenzie & Podsakoff 2012; Podsakoff et al. 2003, 2012). Some of them come from the characteristics of the raters from which the measures are obtained, while others come from item characteristics or the context in which the items are measured and still others come from the measurement context (**Figure 5**). **Table 2** defines the various types of method sources that fall into each of these categories and presents evidence (or examples) of their effects.

When Is Common Method Bias Likely to Be Most Problematic?

The complexity arising from the fact that multiple sources of CMB are likely to be present in any study is compounded by the fact that a respondent's behavior when completing a questionnaire is a function of several factors, including the respondent's ability, experience, motivation, and opportunity to exert less effort (i.e., satisfice) as well as the difficulty of the task (Krosnick 1991, 1999; MacKenzie & Podsakoff 2012; Podsakoff et al. 2012). **Table 3** combines the various sources of CMB with the circumstances under which these biases are likely to cause the biggest problems. For example, rating source is likely to be a problem when respondents lack the ability or the necessary experience with the topics included on the questionnaire, lack motivation (e.g., because of low personal relevance of the issue, low need for cognition, low need for self-expression, the

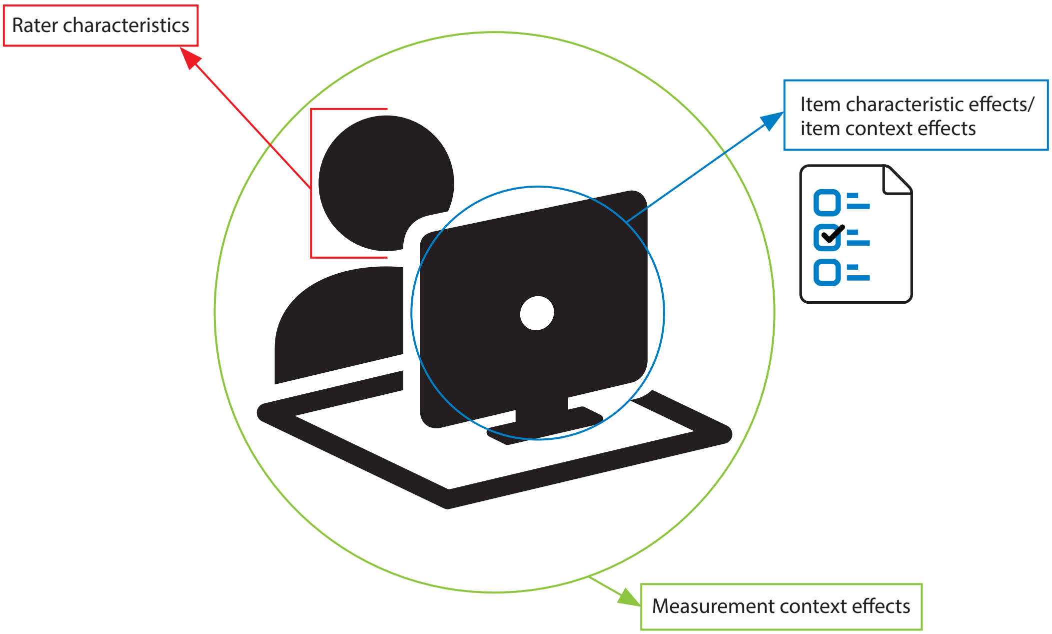

Figure 5

Illustration of the primary sources of common method bias.

presence of implicit theories), or are given the opportunity (or have the motivation) to satisfice in responding. On the other hand, item characteristics and item context effects are likely to be problems when respondents lack motivation (because the repetitiveness of items or the scale is too long), the questionnaire is too difficult (e.g., questions are complex, abstract, or ambiguous or require retrospective recall), or respondents are given the opportunity or the motivation to satisfice (e.g., because scales contain common properties, items measuring the same constructs are grouped together, or answers to similar questions are in close proximity).

Finally, measurement context is a concern when respondents lack the motivation to respond (e.g., because the context arouses suspicions about the researcher's intent, the source of the survey is disliked or distrusted) or because the task is too difficult (e.g., the survey is conducted over the phone and does not allow the respondent to read the questions). Therefore, researchers interested in controlling for CMB need to pay attention not only to the various sources of CMB but also to factors such as the respondents' ability and motivation and the difficulty of the survey, which are likely to heighten the effects of these biases when they are present.

COMMON METHOD BIAS IS WIDESPREAD

Although there is ample evidence of the harmful effects of CMB, some might argue that the conditions necessary for these biases are not prevalent. Unfortunately, that is not accurate. Indeed, there is substantial evidence of the susceptibility of empirical relationships to CMB in several research disciplines, including applied, occupational, positive, and general psychology (Ackerman et al. 2018, Bodner 2006, Pujol-Cols & Lazzaro-Salazar 2021, Sackett & Larson 1990, Spector & Pindek 2016); organizational and vocational behavior (Cooper et al. 2020, Crampton & Wagner 1994); career management (Modem et al. 2022); general, international, and operations management (Chang et al. 2010, Craighead et al. 2011, Fuller et al. 2016); public administration

Table 2 Summary of potential sources of method bias

Potential cause	Definition	Examples/evidence of effects
Common rater effects	Covariation observed between items (or latent variables) on a questionnaire produced by the fact that the same respondent provides the measures of these items (or variables)	
Implicit theories	A priori beliefs that respondents have about the relationships among attitudes, values, perceptions, behaviors, etc.	Research indicates that respondents may believe that certain leadership traits or behaviors (e.g., charismatic leadership behaviors) are related to leadership effectiveness, leading them to rate leaders that they perceive as high on charisma to also be high on effectiveness, even though that may not be accurate (Eden & Leviatan 1975, Lord et al. 1978). Implicit theories have also been used to help explain the relationships between job satisfaction and job performance (Smither et al. 1989), attributions of the causes of group performance (Bachrach et al. 2001, Staw 1975), employee silence (Detert & Edmondson 2011), and OCBs and performance evaluations (Podsakoff et al. 2013).
Consistency motif	Tendency for respondents to try to maintain consistency in their responses to similar items on a questionnaire or to organize their responses in a consistent manner	Respondents asked to rate items that reflect similar content areas (either within or between constructs) will try to maintain consistency in their ratings, thereby inflating the covariation between these items or constructs (Podsakoff & Organ 1986, Schmitt 1994). These biases may be particularly likely to occur when respondents are asked to provide retrospective accounts of their attitudes, behaviors, and perceptions (Podsakoff et al. 2003).
Social desirability	Tendency for respondents to respond to items in a way that puts them in a favorable light or that is viewed favorably by others, rather than on the basis of their real feelings	The correlation between one variable and another may be influenced by the respondents' desire to be viewed in a positive light. Among the topics that have typically been shown to be influenced by social desirability are self-reports of personality traits (Edwards 1957) and feelings of self-esteem (Astra & Singg 2000, Greenwald & Farnham 2000, Huang 2013). However, social desirability has generally not been shown to be related to job performance (Ones et al. 1996).
Leniency biases	Tendency for respondents to provide ratings of themselves or others that are more favorable than is warranted	Respondents are likely to rate the personality, attitudes, beliefs, and intentions of people they like more favorably than those of people that they dislike. In addition, Cheng et al. (2017) report that a rater's personality is related to leniency biases; they found that agreeableness and extroversion are positively related to rating other people leniently and that agreeableness and conscientiousness are positively related to self-ratings, whereas neuroticism is negatively related to self-ratings.
Response styles	Tendency for respondents to systematically differ in the use of response scales Common response styles include the tendency to acquiesce (agree) or to disacquiesce (disagree) with items, irrespective of the content of the item, or to use extreme, midpoint, or noncontingent scale points	Baumgartner & Steenkamp (2001) found that five different response styles (acquiescent, disacquiescent, extreme, midpoint, and noncontingent response styles) accounted for 27% of the variance in the magnitude of the correlations between 14 consumer constructs. They also found that the correlation between constructs can be biased upward or downward depending on the correlation between the response styles. Moreover, Weijters et al. (2010) found evidence that acquiescent and extreme response styles are largely consistent over the course of a survey. To the extent that these tendencies account for systematic variance in the relationships between variables that is different from the true score (actual) variance that exists between these variables, it is problematic.

(*Continued*)

Table 2 *(Continued)*

Potential cause	Definition	Examples/evidence of effects
Positive/negative affectivity, emotionality, or mood	Tendency for respondents to view themselves or the world around them in positive emotional terms (e.g., enthusiasm, joy, energy, cheerfulness) or in negative emotional terms (e.g., fatigue, sadness, disgust, distress)	Several researchers (Connolly & Viswesvaran 2000, Thoresen et al. 2003) have noted that the covariation between items (or constructs) on a questionnaire may be influenced by respondents' tendencies to view the world in generally positive (or negative) terms, irrespective of the content of the items. To the extent that these tendencies account for systematic variance in the relationships between variables that is different from the true score (actual) variance that exists between these variables, it is problematic.
Transient mood state	Tendency for the moods, feelings, or mental states of respondents to be influenced by recent events that they have experienced	The covariation between variables may be influenced by the fact that events experienced by respondents recently (e.g., positive feedback from a supervisor or peer, news suggesting that a global pandemic is worsening, death of a close friend, receiving word of a salary increase) may subsequently put them in a particular mood that influences their responses to items on the questionnaire (Podsakoff et al. 2003). Meta-analytic findings (Lench et al. 2011) indicate that (*a*) transient mood states brought on by emotion elicitation techniques produce correlated changes in behavior, physiology, and experience, and (*b*) elicitation of happiness tends to have stronger effects on outcomes than elicitation of negative emotions (e.g., sadness, anger, anxiety).
Item characteristic effects	Covariation observed between items (or variables) on a questionnaire caused by the specific properties or characteristics that the items (variables) possess	
Item wording	Items (constructs) worded in such a way as to emphasize negative (versus positive) connotations may influence the observed relationships between constructs.	Harris & Bladen (1994) report that the average correlation between stress-related constructs (role ambiguity, role conflict, role overload, job tension) and job satisfaction increased by 238% (from 0.21 when item wording effects were controlled to 0.50 when item wording effects were not controlled) and that the effect of this bias varied depending upon the constructs examined.
Item social desirability	Items written in such a way as to reflect more socially desirable attitudes, values, beliefs, traits, behaviors, or perceptions	To the extent that measures of both the predictor and criterion variables possess social desirability, the covariation between these variables may be different from their true relationships (Podsakoff et al. 2003). Edwards (1953) reported that judges' ratings of the social desirability of items contained in the MMPI correlated 0.87 with the probability that subjects would endorse these items. Similar results were reported by Thomas & Kilmann (1975) for two different scales measuring five different conflict handling modes and by Chen et al. (1997) for measures of positive affectivity and negative affectivity. Finally, across a series of studies, Cui et al. (2022) found that both self-ratings and peer ratings of a variety of personality variables were equally susceptible to item social desirability.
Item demand characteristics	Items written in such a way that they provide cues to respondents that signal how they are expected to respond to them	To the extent that different items (constructs) contain demand characteristics, this may influence the covariation between these variables.

(*Continued*)

Table 2 (*Continued*)

Potential cause	Definition	Examples/evidence of effects
Item ambiguity	Items written in a vague, unclear, confusing, or imprecise manner such that their meaning is ambiguous to respondents	Respondents asked to rate ambiguous items (or constructs) on a questionnaire may respond to these items (constructs) using their own personal heuristic; to the extent that they use the same heuristic across multiple items (constructs), this may influence the covariation between these variables.
Common scale formats	Items on a questionnaire written with the same scale format (e.g., agreement, frequency, similarity)	Podsakoff et al. (2013) argue that respondents exposed to the same scale formats may be less motivated to exert the cognitive energy necessary to process the information contained in the question and be more likely to exhibit undifferentiated responses, subsequently increasing the consistency of responses across the survey items and the likelihood of method biases. However, neither these authors nor Spector & Nixon (2019) found evidence that common scale formats produce stronger relationships between constructs than different scale formats.
Common scale anchors	Items on a questionnaire written with the same scale anchors (e.g., "strongly disagree" to "strongly agree") or the same number of anchor points	Podsakoff et al. (2013) argue that repeated exposure to the same scale anchors decreases respondents' motivation to exert the cognitive effort necessary to process the information contained in scale items and increases the probability of undifferentiated responses, which subsequently increases the consistency across scale items and the likelihood of method biases. Consistent with this explanation, these authors reported that estimates of the relationship between OCBs and performance evaluations were 39% larger when studies used the same (versus a different) number of anchor points when assessing both constructs.
Positively and negatively worded items	Items on a questionnaire written with the same evaluative (positive or negative) wording	Several studies (Greenberger et al. 2003, Harvey et al. 1985, Ibrahim 2001) have demonstrated that negatively worded items often produce method factors that are composed solely of negatively worded items, raising concerns about the construct validity of the measures. Schmitt & Stults (1986) note that these effects may result from the fact that once respondents establish a pattern of responding to survey items, they may ignore the positive–negative wording of the items.
Item context effects	Covariation observed between items (or variables) produced by their relationship to one another on a questionnaire	
Item priming effects	The placement of items (constructs) on a questionnaire may make subsequent items (constructs) on the questionnaire more salient and imply a relationship between the variables.	Janiszewski & Wyer (2014) have noted that the exposure to an initially encountered stimulus (item) makes the processing of that stimulus more accessible and may subsequently influence all stages of the survey response process, including attention, comprehension, memory retrieval, inference, and response generation. Accordingly, researchers (Judd et al. 1991, Tourangeau et al. 1991) have reported that prior questionnaire items affect the speed with which respondents answer similar subsequent questions and that the size of this effect is a function of how closely related the subsequent items are to the initial item prime.

(*Continued*)

Table 2 (*Continued*)

Potential cause	Definition	Examples/evidence of effects
Item embeddedness effects	Items embedded among other positively or negatively worded items will take on the evaluative properties of those items.	Harrison & McLaughlin (1993) reported that evaluatively neutral items that were positioned in blocks of positive (or negative) evaluative items were rated similarly to the items they were embedded in. In a follow-up study, Harrison et al. (1996) reported that the correlation between subjects' ratings of outcome favorability and perceptions of fairness was only 0.10 in a positive measurement context but increased to 0.50 in a negative measurement context, and these correlations differed significantly.
Grammatical redundancy	Items are grammatically redundant when they are identical except for slight variations in item wording.	Cortina et al. (2020) have noted that scales that possess grammatical redundancy (as opposed to conceptual redundancy) are likely to cause respondents to respond consistently and inflate the relationships among items of a construct. Given the similarity in the wording of items that has been shown across some constructs (Spector et al. 2010), it is possible that grammatical redundancy may inflate the covariation across constructs as well.
Item proximity	Items positioned within close proximity to one another, regardless of content	Weijters et al. (2009) reported that the correlations between items measuring unrelated constructs increased by 225% (from 0.04 to 0.09) when they were positioned next to one another than when they were positioned six items apart.
Context-induced mood	Items appearing early on a questionnaire that elicit a positive or negative mood may evoke that mood for the remainder of the questionnaire.	Research from hundreds of studies examining affect induction techniques (Joseph et al. 2020) indicates that a variety of stimuli (e.g., recall of autobiographical events, viewing photos or videos, imagining hypothetical scenarios, body positioning or posture) can elicit emotions. As noted above, transient affective states can bias responses to questionnaire items and other types of measures.
Scale length	Questionnaires that contain fewer items make it easier for respondents to recall responses to previous items from short-term memory. Scales that are excessively long may cause fatigue and motivate respondents to expend less energy in responding to questions (satisfice).	Several authors (Harrison et al. 1996, Podsakoff et al. 2003) have noted that short scales increase respondents' ability to access previous responses and therefore may inflate correlations among items. Moreover, Cortina et al. (2020) have noted that, even though scales have tended to get substantially shorter over the past quarter-century, the internal consistency and reliability (measured by Cronbach's alpha) have tended to remain about the same. They attribute these findings, in part, to the inflation that results from scales that possess higher levels of grammatical redundancy.
Grouping of items (or constructs) on questionnaire	Items from different constructs that are grouped together may decrease intraconstruct correlations and increase interconstruct correlations. In contrast, items from the same construct grouped together may increase intraconstruct correlations and decrease interconstruct correlations.	Schriesheim (1981a,b) indicates that grouping items measuring the same dimensions together produces more leniency response bias than measuring the dimensions with items that have been randomly distributed throughout the questionnaire.Weijters et al. (2009) show that positioning items measuring unrelated constructs next to one another (similar to item grouping) increases the correlation between the items by 225% when compared with positioning the items six items apart.

(*Continued*)

Table 2 *(Continued)*

Potential cause	Definition	Examples/evidence of effects
Measurement context effects	Covariation observed between items (or variables) produced by the physical or psychological context in which the measures are obtained	
Predictor and criterion variables measured at the same point in time	Items measuring the predictor and criterion variables may be obtained from respondents concurrently.	Podsakoff et al. (2003, p. 885) have noted that data gathered at the same (as opposed to a different) point in time "may (*a*) increase the likelihood that responses to measures of the predictor and criterion variables will coexist in short-term memory, (*b*) provide contextual cues for retrieval of information from long-term memory, and (*c*) facilitate the use of implicit theories when they exist." This observation suggests that relationships among items (or constructs) measured at the same point in time will be stronger than when measured at different points in time.
Predictor and criterion variables measured in the same location	Items measuring the predictor and criterion variables may be obtained from respondents in the same location.	Responses obtained in the same location reduce the likelihood of differential cues in the environment that might cause distractions for respondents, thereby strengthening the relationships between the measures.
Predictor and criterion variables measured using the same medium	Items measuring the predictor and criterion variables may be obtained from respondents using the same medium (paper and pencil, computer screen, interview, etc.).	Responses obtained in the same medium reduce the likelihood of differential cues that might cause distractions for respondents, thereby strengthening the relationships between the measures.

Abbreviations: MMPI, Minnesota Multiphasic Personality Inventory; OCB, organizational citizenship behavior.

Supplemental Material >

(Jakobsen & Jensen 2015, Meier & O'Toole 2013); tourism, hospitality, supply chain, and sports management (Kaltsonoudi et al. 2021, Kaufmann & Saw 2014, Kock et al. 2021, Min et al. 2016, Montabon et al. 2018, Zhu et al. 2022); marketing (Hulland et al. 2018); and management information systems (Cram et al. 2019) (for a summary, see **Supplemental Appendix D**). These studies examined almost 13,000 articles and found that between 31% and 98% (and, on average, almost 70%) of the studies reported in the articles published across these disciplines are potentially susceptible to the effects of CMB, either because they obtained the focal variables from the same source at the same point in time or because the studies were susceptible to one or more other sources of CMB. Thus, it appears that researchers in a variety of disciplines have good reason to be concerned about the potential effects of CMB.

COMMON METHOD BIAS IS NOT EASY TO FIX

Given that (*a*) there are various sources of CMB, (*b*) several of these sources are likely to be present in studies using questionnaires to obtain measures of the focal variables (e.g., predictors, mediators, moderators, criterion variables), (*c*) each of the sources may require a different treatment to control its effects, (*d*) different constructs and measures may be susceptible to different method factors, and (*e*) the procedures used to control some biases may exacerbate the effects of others (e.g., reducing the length of a scale to reduce the motivation to satisfice makes it easier to recall previous questions from memory), it is important to understand not only the remedies that researchers have typically used to control these biases but also their efficacy. Therefore, in this section we first identify the procedural and statistical remedies that have been most frequently used to address CMB. Next, we discuss the strengths and limitations of these remedies. We highlight some of the most widely used statistical remedies because, although they rely on questionable techniques, they have been

Table 3 Summary of the potential causes of common method bias and circumstances under which it is likely to be a problem

Causes associated with	Lack of ability, education, or experience	Lack of motivation to respond accurately	Task (questionnaire) is too difficult	Opportunity or motivation to satisfice
Rating source	Lack of verbal ability, education, or cognitive sophistication Lack of experience thinking about the topic/issue being addressed	Low personal relevance of the issue Low self-efficacy to provide a correct answer Low need for cognition Low need for self-expression Low feelings of altruism Low levels of agreeableness Impulsiveness Dogmatism, rigidity, or intolerance of ambiguity Implicit theories Forced participation	Not applicable	The availability of answers to previous questions (in memory)
Item characteristics or item content	Not applicable	Repetitiveness of items Lengthy scales	Complex or abstract questions Item ambiguity Double-barreled questions Questions that rely on retrospective recall Items with long extended time referents (e.g., past 6 months or 1 year)	Common scale attributes (e.g., the same scale types, scale points, and anchor points) Common wording/ grammatical redundancy Grouping related items together The availability of answers to previous questions (due to close physical proximity)
Measurement context	Not applicable	Contexts that arouse suspicions Measurement conditions that make the consequences of a response salient Source of survey is disliked Presence of an interviewer	Auditory-only presentation of items (telephone) versus written presentation of items (print or Web)	The questionnaire is administered using the same medium, in the same location, at the same time.

Table adapted with permission from MacKenzie & Podsakoff (2012) and Podsakoff et al. (2012).

cited (Bozionelos & Simmering 2022, Cruz 2022, Fuller et al. 2016) as support for claims that CMB is generally not a problem in the organizational sciences.

What Are the Most Common Remedies for Dealing with Method Bias?

Almost a dozen studies have reported on the statistical and procedural techniques used by researchers to control for CMB (see **Supplemental Appendix E**). Although these studies vary considerably in the amount of detail they provide, several points are worth noting. First, it is obvious that some academic fields devote more attention to, and are more aware of, the potential harmful effects of CMB than others. For example, whereas between 59% and 81% of articles in

Supplemental Material >

the fields of human resources management (59.1%; Bozionelos & Simmering 2022), international marketing (61.5%; Baumgartner & Weijters 2021), and vocational behavior (80.6%; Cooper et al. 2020) mention the potential effects of CMB in their studies, fewer than 40% of articles in the fields of general business (39.4%; Fuller et al. 2016), supply chain management (34%; Montabon et al. 2018), and information systems (29%; Aguirre-Urreta & Hu 2019) and fewer than a quarter of articles in sports management (21.2%; Kaltsonoudi et al. 2021) and tourism management (13.7%; Kock et al. 2021) mention CMB as a potential threat to the validity of their findings. Second, the most-used post hoc statistical remedy is Harman's single-factor test (HSF), either by itself or in combination with other statistical procedures. The use of HSF is followed by some version of the marker variable (MV) technique (e.g., a correlation-, regression-, or CFA-based marker) and the unmeasured latent variable (UMLV) technique. Few studies use the directly measured latent variable (DMLV) technique or the instrumental variable technique or attempt to control for response styles. Third, although procedural remedies tend to be used somewhat less often than statistical remedies in most disciplines (for an exception, see Bozionelos & Simmering 2022), obtaining measures of the focal variables from different sources and using temporal separation are the most common procedural techniques, followed by proximal or psychological separation or attempts to mitigate biases associated with similar item characteristics (e.g., same versus different scale type, scale anchor points). Finally, Bozionelos & Simmering (2022) reported that almost 7% of the studies they examined in the human resources domain used an unknown test and another 3.5% simply examined the correlation matrix to determine the impact of CMB. Thus, more than 10% of the studies in this domain did not provide adequate information about the potential effects of CMB.

General Procedural Remedies for Controlling Common Method Bias

The four basic procedural remedies for dealing with CMB are obtaining measures of the focal variables from different sources (**Figure 6*a***), introducing a (temporal, psychological, or proximal) separation between the measures of the focal variables (**Figure 6*b***), protecting respondents' anonymity and reducing evaluation apprehension (**Figure 6*c***), and minimizing the common scale properties of the focal variables (**Figure 6*d***). As noted by Podsakoff et al. (2003), the key to using these remedies is to identify what the measures of the focal variables have in common and to remove or minimize these common method features through the design of a study.

Obtaining measures from different sources. Obtaining measures of the predictor and criterion variables from different sources (e.g., other people, objective measures, archival data) breaks the connection between the measures of these constructs that were observed to be "correlated" due to common rater effects (**Figure 6*a***). The objectives of this technique are (*a*) to decrease biases associated with implicit theories, consistency motifs, and transient mood states; (*b*) to reduce tendencies to respond in a socially desirable or lenient manner across the measures of the focal constructs; and (*c*) to minimize the effects of gathering the data at the same time in the same location using the same medium (Podsakoff et al. 2003, 2012). Evidence of the effectiveness of this remedy is depicted in **Figure 4**, which indicates that gathering measures of focal variables from a different (as opposed to the same) source reduces the correlation (on average) between these variables to between 160% and 250%. Thus, when appropriate, obtaining measures of the focal variables from different sources is an effective remedy to the effects of CMB.

Traditionally, alternatives to self-reports include objective indicators (e.g., number of units produced, sales dollars, percentage of sales quota, number of days absent, voluntary turnover) and ratings from other individuals (e.g., supervisors, peers, customers, or on occasion a spouse or a

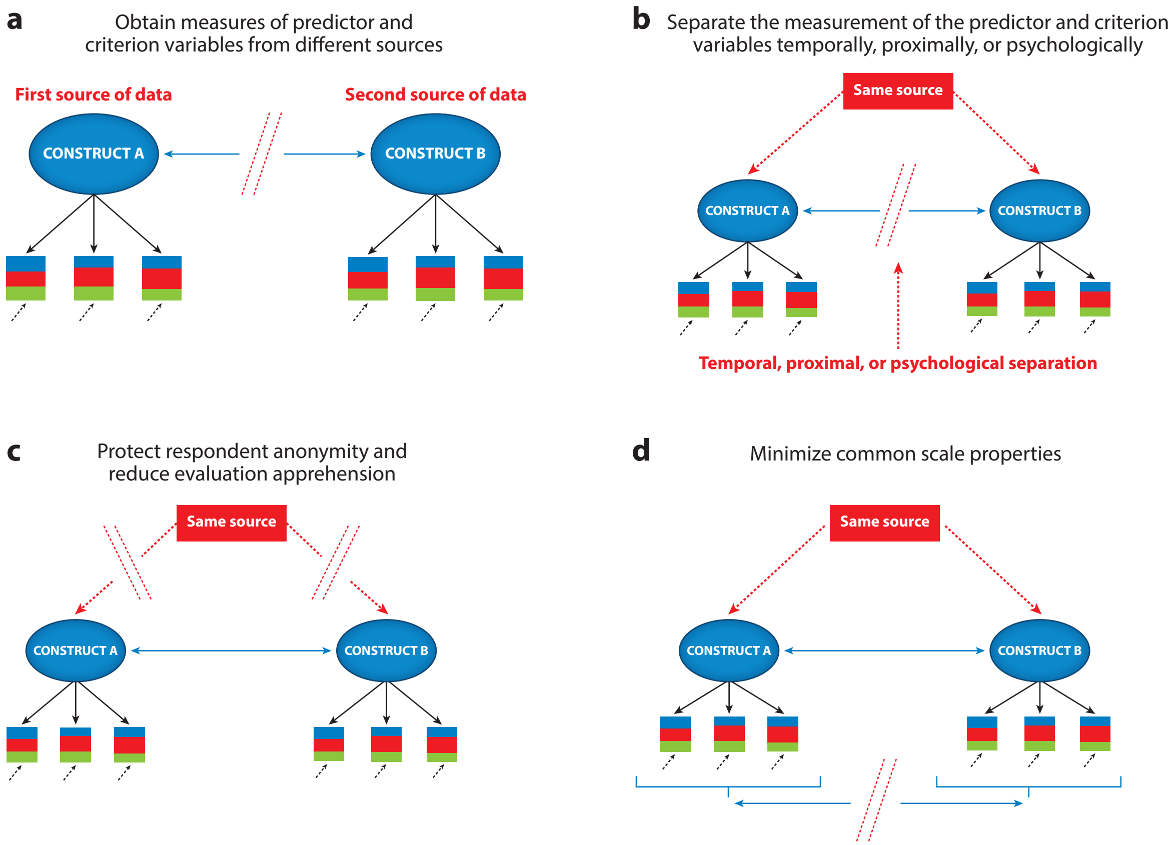

Figure 6

Illustrations of the basic types of procedural remedies. (*a*) Obtaining measures of predictor and criterion variables from different sources. (*b*) Separating the measures of the predictor and criterion variables temporally, proximally, or psychologically. (*c*) Protecting respondent anonymity and reducing evaluation apprehension. (*d*) Minimizing common scale properties. The dotted red lines represent the broken connections between measures of two constructs that were observed to be "correlated" due to common rater effects.

significant other). For example, in their study of the moderating effect of long-term orientation on the relationship between corporate social responsibility (CSR) and new venture financial performance, Wang & Bansal (2012) measured long-term orientation and financial performance from CEOs and presidents of new ventures and CSR from the firms' websites. As another example, Boehm et al. (2014) gathered measures from six different sources to examine the effects of age-inclusive human resource practices on the development of an organization-wide diversity climate, collective perceptions of social exchange, firm performance, and employees' collective turnover intentions. However, researchers have begun to use techniques employed in other disciplines to go beyond the use of self-report measures. For example, Ganster & Rosen (2013) reviewed the use of physiological measures from the field of medicine (e.g., blood pressure, cholesterol, and body mass index) to assess employee well-being, and Chaffin et al. (2017) explored the utility of wearable sensors to obtain measures of behavioral constructs at both the individual and group levels. Although Chaffin et al. caution researchers to be aware of the potential sources of error when using these sensors, they also highlight the possible benefits of working together with the developers of these devices for the purpose of extending our current toolkit. We agree and encourage researchers to think about ways that emerging technologies can be used to gather data from other sources that reduce the potential biases associated with self-reports.

Of course, acquiring measures of the focal variables from different sources is not without limitations. For example, it is difficult to use this procedure when focal variables represent an individual's internal states (e.g., attitudes, beliefs, values, intentions) because (*a*) obtaining valid measures of a person's internal states by others requires them to accurately infer these events from the person's behavior, even though (*b*) the individual's internal states may not translate into observable behaviors and (*c*) observers may not have the ability or opportunity to observe these behaviors even if they do occur (Brannick et al. 2010, question 7; Podsakoff et al. 2012). Moreover, obtaining objective data such as job performance is difficult because some jobs (e.g., white collar and managerial jobs) may not have readily available objective measures (Hopp et al. 2009), objective indicators may capture only some of the important aspects of overall job performance (Rotundo & Sackett 2002), and organizations may be unwilling to share some forms of archival data (e.g., personnel records) because of legal or other constraints. Finally, obtaining data from different sources and matching the information from the predictor and criterion variables generally take more time, energy, and resources than does obtaining data from only one source.

A variation of this remedy that is often used in research on groups and teams is to randomly split a sample of employees from a single work unit into two groups and then to gather the predictor variable(s) from one of the subsamples and the criterion variable(s) from the other subsample. This split-sample design has been operationalized in different ways. Some studies (e.g., Smith et al. 1983, Zohar & Luria 2004) split their overall sample to obtain aggregated measures of the predictor variables from the first subsample and obtain aggregated measures of the criterion variables from the second subsample. In doing so, they lessened the impact of same-source biases on the estimates of the substantive relationships between predictors and criteria. Alternatively, the split-sample technique can be applied to the measurement of multidimensional constructs to remove same-source biases from the ratings of different dimensions of the same construct. For example, Schulte et al. (2009) split participants of each unit into seven groups and collected ratings of one of seven different climate subdimensions from each group to alleviate concerns that response bias could inflate the interdimensional relationships.

Although splitting the sample using either of these techniques has the advantage of gathering the predictor and criterion variables from different sources, these remedies are not without limitations. First, as noted by several authors (Bliese 2000, Morgeson & Hofmann 1999), researchers should not assume without good theoretical reasons that the structure or function of individual-level constructs generalizes to the group or organizational level (i.e., that the constructs are isomorphic). Second, it is important for the referent of the measures to reflect the level of the construct (individual or group) being operationalized (Chan 1998, Podsakoff et al. 2015). Third, with some composition models, the aggregation of the focal variables must be justified statistically by demonstrating that a substantial, meaningful amount of variance in measures of the focal variables is attributable to between-group factors (Bliese 2000, Chan 1998). Fourth, aggregating both the predictor and criterion variables will reduce the sample size and the subsequent power of the statistical test used, thereby increasing the likelihood of Type II errors (Podsakoff & Organ 1986). Finally, using the responses of coworkers as surrogates for individual-level ratings of some constructs is questionable. For example, aggregating coworker responses of their leaders' behavior is inconsistent with research showing that leaders respond differently to different employees (Henderson et al. 2009, Martin et al. 2018). However, assuming that there are no theoretical or methodological restrictions from aggregating the data, these remedies do have advantages for those concerned with CMB associated with gathering data from the same source.

Introducing a separation between the predictor and criterion variables. When it is impossible or inappropriate to obtain measures of the focal variables from different sources, another procedural remedy that aims to reduce CMB is to introduce a separation between the predictor and criterion variables included in the study (**Figure *6b***). Although this technique acquires measures of the focal variables from the same source, it separates these measures temporally, psychologically, or proximally. This remedy may prove particularly useful when examining the relationships between internal states (e.g., attitudes, beliefs, moods, values, perceptions, intentions) or behaviors that are difficult for other individuals to observe because they have a low base rate or are purposely hidden from view (e.g., sabotage, retaliation, unethical behavior).

Our review of meta-analytic studies that examined the effects of temporal separation between the predictor and criterion variables (**Figure 4**) indicates that, except for positive–negative valence pairs, gathering measures of these variables at different points (versus the same point) in time decreases the correlation (on average) between 121% and 142%. These findings suggest that another way of reducing CMB is to introduce a temporal separation between the focal variables. Examples of the use of temporal separation include studies conducted by Gielnik et al. (2018), who examined the moderating role of age-related factors (future time perspective and prior entrepreneurial experience) on the relationships among entrepreneurial identification, entrepreneurial intentions, and self-reported entrepreneurial activity in a three-wave study, and by Loi et al. (2020, study 1), who examined the sequential mediating effects of moral licensing (moral credits and moral credentials) and psychological entitlement on the relationships between employee volunteering and workplace deviance in a four-wave study.

However, this remedy has limitations. First, since we lack a comprehensive understanding about the length of time over which predictor variables have their effects in our field (Mitchell & James 2001, Shipp & Cole 2015, Shipp & Jansen 2021), the lags we use in our studies may be either too short (thus proving ineffective) or too long (thereby allowing other factors to influence the outcome variables or the focal effect to dissipate) (Podsakoff et al. 2012, Spector 2019). This lack of understanding is compounded by the fact that different phenomena are likely to require different temporal delays. Second, in our role as reviewers, we have encountered studies that employ time delays as short as 1 h. This short duration is problematic, because research (Johnson et al. 2011, Ostroff et al. 2002) has shown that although 3-week to 1-month delays can cause an appreciable reduction in the relationships between focal variables, a 1-h delay does not. Third, implementing a temporal delay assumes that the relationships between the predictor and criterion variables are stable over time; otherwise, conclusions about the dissipating effects of CMB may be incorrect. This may be particularly important in the case of some variables, such as employee affect (Beal et al. 2005), moods (Scott et al. 2020), and emotions (Lim et al. 2018), which fluctuate considerably over time (Podsakoff et al. 2019). Moreover, gathering data over time may result in respondent attrition and is likely to take additional time and energy on the part of the research team. Finally, although this remedy should help minimize the potential effects of some of the sources of bias related to raters, including consistency motifs and transient mood states, as well as some of the item-related sources of bias, it is unclear whether it counters the effects of leniency biases, response styles, or trait social desirability. That may be why our analysis (**Figure 4**) found that temporal separation had less of an effect on the relationships between focal variables than did the use of measures from different sources. Despite these limitations, temporal separation is likely not only to enhance the causal inferences between the predictor and criterion variables but also to assuage concerns about the effects of CMB.

Of course, a particularly effective way of controlling for CMB is to obtain measures of the focal variables from different sources and include a temporal separation between them. For example, Sessions et al. (2020, study 1) examined the positive and negative effects of employee promotive and prohibitive voice behaviors on supervisor emotional exhaustion and performance. These authors gathered measures of (*a*) employee-reported group promotive and prohibitive voice behaviors (time 1), (*b*) supervisor-reported measures of personal power as well as challenge and hindrance appraisals of group voice behavior (time 2), and (*c*) supervisor-reported emotional exhaustion and subordinate-reported supervisory behavior (time 3), with each time period lagged by approximately 1 month. Other examples of studies that incorporate measures of the focal variables from different sources over different time periods include those by Nifadkar et al. (2012, study 2), who examined the mediating effects of supervisor-triggered newcomer affect on the relationships between supervisory behaviors and newcomer behavior, adjustment, and performance, and by Deng et al. (2021, studies 1 and 2).

Although conducting research that includes data from multiple sources and multiple time periods is not easy, it is obvious that the strengths of these studies are that they reduce the possibility that various sources of CMB influence the results and that they increase causal inferences. Thus, we strongly encourage future research using this combination of remedies.

An interesting experimental study by Rubenstein and colleagues (A.L. Rubenstein, personal communication, Sept. 6, 2022) presents evidence of the effects of psychological separation. These authors examined the effects of a task administered midway through a survey that was intended to divert respondents' attention away from previously accessed answers and disrupt cognitive consistency on correlations across a variety of constructs commonly examined in applied psychology. These constructs included job attitudes, job performance, personality, aptitude, and demographics/biodata. The authors tested a psychological separation of 5 min, 7.5 min, and 10 min. The criterion they used was that the psychological separation task had to produce the same reduction in the size of the correlations among the variables of interest that they obtained using different sources for the measurement of the predictor and criterion variables, or a temporal separation of 2 weeks. Of the 39 relationships examined in their study, Rubenstein and colleagues found that none met the criteria for the 5-min psychological separation task. However, 20 of them met the criteria for the 7.5-min separation task and 12 of them met it for the 10-min separation task, although the number of relationships that met the criteria depended on the type of the variables examined. More specifically, in the case of the 7.5-min separation task, (*a*) all attitude–attitude correlations (four of four) met the criteria, (*b*) all eight of the attitude–performance and performance–performance correlations met the criteria, (*c*) the majority (five of seven) of the personality–attitude correlations met the criteria, and (*d*) a minority (three of eight) of the personality–performance correlations met the criteria. However, none of the five relationships between aptitudes and employee attitudes or between personality and job performance, nor any of the six personality–personality correlations, met the criteria. Taken together, these findings indicate that psychological separation may prove to be a valuable tool for controlling CMB for some (but not other) types of relationships—particularly when the relationships being examined include attitude–attitude, performance–performance, attitude–performance, and personality–attitude construct pairs.

With that said, we believe that it is premature to generalize these findings because (*a*) this is the first study to examine the effects of psychological separation on the relationships between variables, (*b*) the finding that a 7.5-min psychological separation is more efficacious in reducing the relationships than the 10-min separation deserves more attention, (*c*) the temporal separation was only 2 weeks long, and (*d*) the results need to be replicated using other combinations of variables. Moreover, given that recent research by Spector et al. (2022) indicates that CMB affects different measures of the same construct differently, and because we know little about the possible effects that the filler task used in the Rubenstein and colleagues study had on the findings, more research needs to be conducted to determine the overall efficacy of this approach. Nevertheless, the preliminary evidence supporting the efficacy of this procedure is encouraging, and we believe it would prove very worthwhile for researchers to conduct additional studies that compare this technique with obtaining measures from different sources and different time waves to determine its generalizability.

Weijters et al. (2009) present evidence of the effects of proximal separation. These authors examined the effects of item proximity and the nature of the conceptual relationship (unrelated, reversed, or nonreversed conceptual meaning) between survey items on the strength of item correlations. They found that the correlation between item pairs measuring unrelated constructs increased by 225% (from 0.04 to 0.09) when these items were positioned next to one another compared with when they were positioned six items apart. In addition, they reported that (*a*) the average correlation between nonreversed item pairs that were positioned six or more items apart increased by 177% (from 0.35 to 0.62) when they were positioned next to one another, and (*b*) the average correlation between reversed item pairs that were positioned six or more items apart decreased by 433% (from −0.26 to −0.06) when they were positioned next to one another. The latter findings suggest that the positive correlation between item pairs becomes weaker for nonreversed items, while the negative correlation for reversed item pairs becomes stronger, the further the items are positioned from one another (Weijters et al. 2009).

Johnson et al. (2011, study 2) also examined the effects of proximal separation by adding filler scales between the measures of their higher-order CSE construct. They found that doing so reduced loadings on the higher-order CSE construct by an average of 12% and decreased the R^2 for the relationships between CSE and job satisfaction by 31%. However, in addition to adding filler scales, these authors changed the response formats of the scales; therefore, it is not possible to isolate the unique effects of proximal separation in their study. Nevertheless, taken together, the studies by Weijters et al. (2009) and Johnson et al. (2011) suggest that researchers who are interested in reducing the effects of CMB may want to consider either dispersing indicators of the same construct throughout the questionnaire, rather than grouping the items together, or adding buffer items between measures of the same construct. Of course, inserting buffer items may reduce the motivation of respondents and increase their fatigue, and using this remedy is likely to reduce factor loadings and reliabilities of the focal constructs. Nevertheless, Weijters et al. (2009) have shown that inserting even a few spaces between items has a nontrivial effect on reducing CMB.

Ensuring respondent anonymity. Another technique that has been used to reduce the effects of CMB is to protect respondents' anonymity (**Figure 6*c***). The purpose of this technique is to reduce the evaluation apprehension that respondents might experience in providing their responses by not asking for personal information that identifies them. Although guaranteeing anonymity should lessen respondents' motivation to edit their responses to be more socially desirable, lenient, and consistent, this procedure is not likely to remove biases based on implicit theories, item characteristics, or item context effects. Moreover, if anonymity is provided for only one, but not both, of the focal variables (i.e., either the predictor or the criterion variable), it makes it difficult if not impossible to match the responses obtained using this technique with data obtained from other sources or over time, unless a linking variable that is not associated with the respondent's identity is used (Podsakoff et al. 2003). Finally, research by Lelkes et al. (2012) suggests that although complete anonymity may reduce a respondent's motivation to distort responses in a socially desirable direction and subsequently increase reports of socially undesirable behavior, it may also decrease accountability, reduce reporting accuracy, and increase satisficing. Therefore, researchers should not routinely assume that providing anonymity to respondents reduces CMB.

Minimizing common scale formats/properties and reducing item social desirability. The fact that several sources of CMB originate from item characteristics and item context effects suggests that minimizing common scale properties and response formats is another way of reducing CMB. **Figure 6*d*** illustrates this remedy with breaks between the items measuring Construct A and those measuring Construct B. Examples include using different scale formats (e.g., agreement, frequency) and scale anchor points, reverse-coding items, and minimizing item social desirability.

As noted by Podsakoff et al. (2013), repeated exposure to the same scale formats and scale anchors decreases respondents' motivation to exert the cognitive effort necessary to process the information contained in scale items, which subsequently increases the consistency across scale items and the likelihood of CMB. Consistent with this explanation, these authors reported that estimates of the relationship between OCB and performance evaluations were 39% larger when studies used the same number (versus a different number) of anchor points. These findings suggest that varying the number of anchor points across measures included in a questionnaire may help reduce CMB associated with item similarity.

The effects of scale formats are not as straightforward. For example, Podsakoff et al. (2013) did not find support for the hypothesis that common scale formats produce stronger relationships than different scale formats, and Spector & Nixon (2019) reported only mixed support for the expectation that scale formats produce differences in the correlations between stress-related constructs. In contrast, both Dalal (2005) and Spector et al. (2010) reported that OCB and CWB are more strongly related when agreement scales rather than frequency scales are used. Thus, additional research on the effects of this methodological procedure is warranted. In any case, note that although the use of some scale formats (e.g., frequency scales) may be appropriate for some constructs (e.g., behaviors), they may be less appropriate for other constructs (e.g., attitudes); thus, researchers also need to ensure that their scale formats match the nature of the focal constructs they are measuring.

Reverse-coding items is designed to minimize acquiescence bias and reduce respondents' tendency to satisfice and respond stylistically by introducing "cognitive speed bumps" that require extra processing time by the respondents. However, there is evidence that this technique may produce spurious factor structures and reduce estimates of construct reliability (Chyung et al. 2018, Weijters & Baumgartner 2012), suggesting that this remedy should be used with caution. On a positive note, Mathews & Shepherd (2002) reported that forewarning respondents about the presence of negatively worded items in the questionnaire reduced the amount of careless responding and the amount of negative factor loadings in their study—but did not eliminate them. In addition, Chyung et al. (2018) noted that grouping together negative and positive items from the same construct on a questionnaire should focus respondents' attention and increase the probability that respondents will process the items more deeply. Nevertheless, researchers should be aware of the potential advantages and disadvantages of introducing negatively worded items into their survey questionnaires.

Finally, several studies (Chen et al. 1997, Cui et al. 2022, Thomas & Kilmann 1975) have shown that judges' ratings of item social desirability are strongly related to the endorsement of these items by survey respondents. In addition, Cui et al. (2022) found that (*a*) self- and peer ratings of personality were equally susceptible to item social desirability and (*b*) the effects of item social desirability were more pronounced when respondents scored high on trait social desirability. These findings suggest that it is important to minimize item social desirability, where possible. This may be accomplished by using neutral (as opposed to socially desirable) items in the questionnaire (Nederhof 1985) or by removing items that correlate highly with scores on a trait social desirability scale (Kam 2013). Note, however, that many constructs in applied psychology and management have positive (job satisfaction, work engagement, helping behavior) or negative (neuroticism, deviant behavior) connotations, and it may be difficult if not impossible to develop valid measures of these constructs while also completely eliminating item social desirability. Moreover, although controlling for item social desirability is possible when developing and validating a scale, this technique is less appropriate when using scales that have already been reported in the literature, because doing so may compromise the validity of these measures. Therefore, although trying to minimize item social desirability in the early stages of scale development is worthwhile, researchers need to temper their desire to control this source of bias if it means compromising the validity of their measures.

Specific Procedural Remedies Where Common Method Bias Is Likely to Be Prevalent

The procedural remedies discussed above attempt to lessen the effects of CMB that stems from rating sources, characteristics of the items or the context in which the items are presented, and/or the measurement context. Although these remedies are valuable for addressing CMB, the fact that such biases are more likely to be present when the ability or motivation of the respondent is lacking, the task (i.e., the questionnaire) is difficult, or the opportunity to satisfice is available (Krosnick 1991, 1999; MacKenzie & Podsakoff 2012; Podsakoff et al. 2012) suggests that procedural remedies focused specifically on these conditions should prove valuable to researchers interested in minimizing the effects of CMB. **Table 4** summarizes the remedies that are tied to these conditions.

Remedies directed at a lack of ability. When researchers are concerned that respondents may lack the requisite cognitive ability or experience in dealing with the topic of interest, they should (*a*) select respondents with the necessary ability, education, and familiarity with the focal topic; (*b*) pretest the questionnaire with participants from the same subject pool; and (*c*) use terminology and grammar that match the respondents' capabilities (**Table 4**). The key here is to ensure that there is a match between the abilities and experiences of the respondents and the questions being asked.

Remedies directed at a lack of motivation. In contrast to a lack of ability, respondents may lack the motivation to respond accurately. Several studies (Anseel et al. 2010, Heberlein & Baumgartner 1978, Roth & BeVier 1998) have shown that survey response rates are significantly higher when the topic being studied is salient to respondents. These findings suggest that researchers may be able to reduce some of the effects of respondents' lack of motivation to respond accurately, or their desire to satisfice, by carefully choosing a sample for whom the topic being studied is interesting, relevant, and/or salient. For example, Colquitt et al. (2011) chose a sample of firefighters to (*a*) distinguish between trust in low-reliability contexts (marked by low levels of unpredictability and danger) and trust in high-reliability contexts (marked by high levels of unpredictability and danger) and (*b*) examine the effects of these contexts on firefighters' physical well-being, withdrawal behaviors, and job performance. In another example, Grant & Hofmann (2011, study 1) chose a sample of university fundraisers to examine the effects of ideological messages from a beneficiary versus organizational leaders on fundraisers' performance. Given the relevance of the issues being examined in these studies to the samples chosen, we doubt that the participants lacked the motivation to complete the surveys. Thus, we encourage researchers to think more carefully about this issue when choosing their sample.

Beyond choosing an appropriate sample, researchers can increase respondents' motivation to respond accurately by (*a*) explaining how answers to the survey have important consequences for them or their organizations, (*b*) emphasizing that their personal experiences are important and that only they can provide them to the researcher, (*c*) explaining how much others (or the organization) are depending on the accuracy of their responses, (*d*) letting them know how much the researcher

Table 4 Specific procedural remedies by CMB source and condition

Source of CMB	Condition to be addressed	Remedies
Rater	Lack of ability, education, or relevant experience with the topic being addressed	Select respondents who have the necessary ability, education, and experience thinking about the topics to be addressed in the study. Match the difficulty of the task with the abilities of the respondents by pretesting the questionnaire with participants drawn from the same sample pool to ensure that the questions are understood. Use language, vocabulary, and syntax that match the reading abilities of the respondents.
	Lack of motivation to respond accurately because respondents do not understand the importance of the information	Choose a sample for which the topic studied is relevant, interesting, and/or salient. Enhance the motivation to answer accurately by explaining how answers to the questions have important consequences for the respondent and/or the organization. Explain how much others (or the organization) are depending on the accuracy of their responses. Emphasize to respondents that their personal experiences are important, and that only they can provide them to the researcher. Enhance the motivation for self-expression by explaining that "we value your feedback," "your opinion is important to us," and/or "we want to know what you think." Treat participants in a respectful manner, let them know you value their time, and express your appreciation for their participation. Promise to provide feedback to respondents to motivate them to respond accurately so that they can gain self-awareness and self-understanding. Personalize the appeal to complete the questionnaire by adding a signature from the researcher.
	Lack of motivation to respond accurately because of suspicions about the measurement context	To alleviate suspicions about who will have access to the data, explain why the information is being gathered and how it will be used, that it is for research purposes only, that only aggregated information will be provided to the organization, and that no individual responses will be given to anyone associated with the organization. Where possible, use anonymity.
	Lack of motivation to respond accurately because of item characteristics or item context	Minimize the repetitiveness of items and their grammatical redundancy. To the extent possible, minimize the length of the questionnaire (without compromising the construct validity of the measures).
Task difficulty	Questionnaire is too complex, ambiguous, or difficult	Use clear, concise language. Simplify complex or abstract concepts and questions. Avoid double-barreled questions. Clarify any vague concepts with examples. Simplify questions and response options.
Opportunity to satisfice	Opportunity to exhibit less cognitive effort (i.e., satisfice) because of item characteristics or item context	Where possible, vary scale types, scale points, and anchor points. Avoid grammatical redundancy in the items. Space related items apart.

Table adapted with permission from MacKenzie & Podsakoff (2012).
Abbreviation: CMB, common method bias.

values their time, (*e*) personalizing the appeal by adding a signature from the researcher, and, where possible, (*f*) promising feedback to participants so that they can gain self-awareness and self-understanding (MacKenzie & Podsakoff 2012).

Of course, in addition to a lack of understanding of the importance of the information being sought, respondents may lack the motivation to expend energy completing the survey either because they are suspicious about the measurement context or because of the characteristics of the items on the questionnaire or the context in which they are solicited. To lessen concerns about the measurement context, researchers can explain (*a*) why the data are being gathered, (*b*) that the data will be used only for research purposes, (*c*) that only aggregated information will be provided to the organization, and (*d*) that individual responses will not be given to anyone in the organization. Moreover, if identifying the respondent is not critical for other reasons (e.g., matching responses from the survey with other information), the responses can be made anonymous. Finally, to reduce the potential negative effects of item characteristics or item context on respondents, researchers should, where possible, minimize the repetitiveness of the items as well as the length of the questionnaire, if doing so does not compromise the construct validity of the measures (MacKenzie & Podsakoff 2012).

Remedies directed at task difficulty. Another condition that has been found to be related to CMB is that the task (survey) is too difficult, which may result from the use of ambiguous, difficult, or complex language. To address these conditions, researchers should (*a*) use clear and concise language, (*b*) simplify complex or abstract concepts and questions, (*c*) clarify vague concepts with examples, (*d*) avoid double-barreled questions, and (*e*) simplify questions and response options.

Remedies directed at item characteristics and item contexts. Finally, given that item characteristics and the context in which they are presented can affect the opportunity to satisfice, researchers interested in reducing the effects of CMB can, where possible, (*a*) vary the types of scales, scale points, and anchor points; (*b*) space related items apart; and (*c*) avoid grammatical redundancy across the measures of the predictor and criterion variables (Cortina et al. 2020). However, when modifying the properties of any scale, it is important to assess the construct validity of the adapted measures (for the types of evidence that can be used to support the validity of adapted scales, see Heggestad et al. 2019).

Statistical Approaches to Detecting and Controlling Common Method Bias

In addition to the procedural remedies discussed above, several statistical remedies may be used to detect and control for CMB. There are good reasons to believe that each of the most popular statistical techniques identified in **Supplemental Appendix E** suffers from important limitations. As latent variable CFA models have come to play a dominant role in CMB research, we focus on these approaches, including HSF, the UMLV technique, the MV technique, and the DMLV technique. All of these techniques can be best understood as extensions of the basic CFA model depicted in **Figure 7*a***. This example shows two substantive correlated factors, each measured by three indicators. The CFA approach to investigating CMB requires the addition of different types of method latent variables to detect and control for the variance resulting from specific measurement techniques used in a given study. These CFA methods can be regarded as extensions of earlier approaches that did not incorporate latent variables, and they are more effective than these earlier approaches because they (*a*) control for random measurement error; (*b*) allow for statistical tests of the presence and impact of CMB; and (*c*) partition the variance of measures into construct, method, and random error variance.

Supplemental Material >

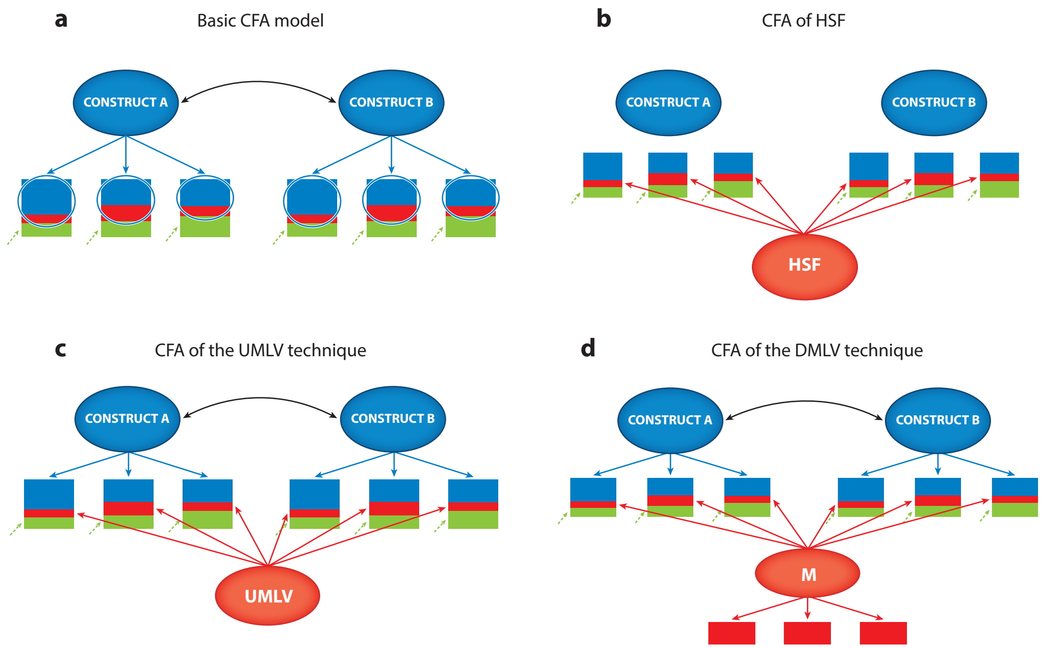

Figure 7

Illustrations of statistical remedies for controlling common method bias. (*a*) Basic CFA model. (*b*) CFA HSF model. (*c*) CFA-based UMLV technique. (*d*) CFA-based DMLV approach. Abbreviations: CFA, confirmatory factor analysis; CMV, common method variance; DMLV, directly measured latent variable; HSF, Harman's single-factor technique; M, the specific method factor used in the analysis (which could include a marker variable, a directly measured latent variable, or measured response styles); UMLV, unmeasured latent variable technique.

Harman's single-factor technique. The assumption of HSF is that if CMV is present and has a confounding effect, it will manifest itself through the presence of a single dominant factor. Historically, the absence of a single dominant factor or the presence of additional factors in a model was used to support the conclusion that CMV was not a problem. In terms of implementation, all the items of the substantive variables included in a study are loaded into an exploratory factor analysis (EFA). If the first unrotated factor does not account for the majority (>50%) of variance in the items, it is inferred that CMB is not a problem. This EFA approach has recently been replaced or supplemented with the CFA HSF model shown in **Figure 7*b***, which excludes substantive latent factors and replaces them with a single method factor that loads on all the substantive indicators. Then the fit of the HSF model is examined using one or more indices for model evaluation, and if it is determined to be worse than the originally proposed substantive model, CMV is not considered to be of concern.

The basic advantage of both EFA and CFA HSF techniques is their ease of implementation. Since they do not require any preplanning at the study design stage, nor do they require researchers to include additional, nonfocal measures in their questionnaires, they are often used in a post hoc fashion as a response to journal editor/reviewer concerns about CMB. However, this convenience is substantially offset by the limitations of HSF. First, as noted by several researchers (Aguirre-Urreta & Hu 2019, Baumgartner et al. 2021, Hulland et al. 2018, Jakobsen & Jensen 2015), the single common factor extracted from EFA is likely to confound substantive variance and CMV, leading to false positives when correlations between substantive constructs are high and false negatives when correlations are medium to low. Relatedly, HSF is subject to low statistical power because it can detect CMV only when the majority of the variance in the ratings of the substantive constructs can be explained by one factor, making the technique very conservative (Aguirre-Urreta & Hu 2019, Baumgartner & Weijters 2021, Baumgartner et al. 2021, Hulland et al. 2018, Jakobsen & Jensen 2015, Schwarz et al. 2017, Steenkamp & Maydeu-Olivares 2021). Third, HSF does not identify the specific sources of method variance affecting the focal relationships, and it assumes a single source of method variance for all the items—which is doubtful, given that more than one source of method variance is likely to be present in all survey studies. Finally, the criterion used to determine how much variance the first factor should explain (e.g., >50%) is arbitrary (Podsakoff et al. 2003, Steenkamp & Maydeu-Olivares 2021).

Although it appears to be more rigorous, the CFA HSF approach suffers because if the hypothesized measurement model fits the data better than a single-factor solution, this procedure provides evidence of the fit of the measurement model and not evidence for the absence of CMB. However, perhaps the biggest limitation of both EFA and CFA HSF techniques is that they do not control for CMB—they simply try to determine whether this form of bias is present in the data by using a very unsophisticated tool. This approach is problematic because even if the first unrotated factor does not account for 50% of the variance in the items included in EFA, the presence of shared method variance across the focal variables will produce biased estimates of the observed relationships. Similarly, the rejection of the CFA-based HSF model does not mean that there is no CMV or bias present that may be compromising researchers' conclusions. Thus, like several other researchers before us (Aguirre-Urreta & Hu 2019; Baumgartner & Weijters 2021; Baumgartner et al. 2021; Hulland et al. 2018; Jakobsen & Jensen 2015; Podsakoff et al. 2003, 2012; Schwarz et al. 2017; Steenkamp & Maydeu-Olivares 2021), we strongly recommend against using either EFA or the CFA HSF technique.

Unmeasured latent variable technique. The CFA-based UMLV technique can be thought of as taking the HSF and adding it to the original correlated substantive factor model (rather than using it to replace the substantive latent variables). **Figure 7*c*** illustrates the resulting model. To test for the presence of CMB, researchers compare (*a*) the fit of the model that includes only the substantive factors and their loadings with (*b*) the fit of the model that includes both the substantive and method variable factors and loadings. If the fit of the two models is not significantly different (e.g., the chi-square difference is less than that of the critical values for its degrees of freedom), then the researcher can conclude that CMV is not present (and infer that CMB is not of concern). However, if the model that includes the unmeasured method factor significantly improves the fit of the model, then it is important to control for these method effects to determine whether the observed relationships between the substantive variables are still significant. If the empirical relationships between the predictor and criterion variables remain significant after the inclusion of the unmeasured latent method factor, researchers can conclude that their relationships hold even after statistically controlling for sources represented by the unmeasured method factors.

The UMLV technique has several potential advantages. First, it is relatively easy to implement after the primary data for the study have been acquired, and it does not require additional measures. Second, the UMLV technique does not require researchers to identify or measure the specific variable(s) responsible for the method effect(s). Third, this technique models the effect of the method factor at the item level, rather than at the construct level (Podsakoff et al. 2003, 2012; Williams et al. 1996). Finally, the UMLV technique does not require the effects of the method factor on each indicator to be equal (Podsakoff et al. 2012).

However, the technique also has several limitations. First, as noted by several researchers (Bagozzi 2011, Podsakoff et al. 2003, Williams & McGonagle 2016), one cannot be sure what specific sources of CMV are being captured by the latent method factor, and the method factor may reflect not only different types of method variance but also variance due to relationships between the constructs other than the one hypothesized. Second, two Monte Carlo simulations (Chin et al. 2012, Richardson et al. 2009) suggest that the UMLV technique does not appear to do an effective job of detecting or controlling for CMB, regardless of whether one uses a traditional covariance-based structural equation model analysis (Williams et al. 1989) or partial least squares (PLS) analysis (Liang et al. 2007). Specifically, on the basis of their simulation, Richardson et al. (2009, pp. 793–94) concluded that the typical covariance-based technique correctly identified "the presence or absence of CMV [only] about 41% of the time" and that such low rates of detection "rarely meet the criteria for usefulness." Similarly, a Monte Carlo simulation of the UMLV technique by Chin et al. (2012, p. 1003) that used PLS led the authors to conclude that it "is neither able to detect, nor control for, common method bias. Method estimates using this approach resulted in negligible estimates, regardless of whether there were some, large, or no method bias introduced in the simulated data." In addition, Chin et al. noted that the PLS-based UMLV technique is problematic because, among other things, every indicator in the PLS model includes not only variance attributable to the trait of interest but also method variance, and any of the structural paths among the constructs will still be biased by method effects.

The UMLV approach suffers from technical challenges as well. For example, Podsakoff et al. (2012) have noted that adding the method factor to the model can cause identification problems if the ratio of the number of indicators to the number of substantive constructs is low. Identification problems also are likely if the method factor loadings have similar or equal values (e.g., Kenny & Kashy 1992). Furthermore, this procedure assumes that the substantive variable factors do not interact with the method factor, an assumption that has been questioned by several researchers (Bagozzi & Yi 1990, Campbell & O'Connell 1967, Podsakoff et al. 2003, Wothke & Browne 1990). However, in our view the biggest problem with the UMLV approach is its use of a single latent variable to represent what are very likely to be multidimensional sources of CMV. This problem was illustrated in a recent study reported by Spector et al. (2019). After applying the UMLV technique to simulated data, these authors concluded that the "results indicate that simplistically modeling one method factor in circumstances where method variance actually stems from multiple method factors with (*a*) distinct patterns of relationships with one another and (*b*) across substantive items can, in some cases, produce more erroneous results than does modeling no method factors at all" (Spector et al. 2019, p. 866).

Marker variable technique. The next set of approaches used for detecting and controlling CMV differ from HSF and the UMLV technique in that they incorporate measures of variables presumed to reflect specific forms of method variance in the analyses. **Figure 7*d*** shows a general model for these approaches. Note that, in contrast to the UMLV, this model includes indicators used to capture the specific types of variance that may be the source of CMV/CMB. The first CFA technique we present uses MVs that are presumed to share measurement characteristics with the substantive constructs of interest but are conceptually unrelated to these variables. Lindell & Whitney (2001) introduced a partial correlation technique for use of MVs that assumes that any empirical relationship between the MV and a substantive variable represents method-related biases. As noted by Richardson et al. (2009), an ideal MV should meet three criteria: It should be (*a*) chosen a priori and (*b*) theoretically unrelated to the substantive variables but (*c*) similar to them in content and format. Although the MV technique was originally designed to control for CMB by partialling out the smallest correlation between the MV and a substantive variable included in the study to determine whether the relationships between the substantive variables were still significant, more recently researchers have used a more sophisticated regression-based analysis (Siemsen et al. 2010). In the regression-based technique, the MV is added to the regression equation, along with the predictors of the focal criterion variables, to determine whether the substantive variables remain significantly different from zero.

Williams et al. (2010) proposed a CFA-based MV technique to overcome the limitations of the partial correlation and regression procedures, which can be understood via **Figure 7*d***. In this approach, a series of measurement models representing the relationships between the latent MV and the substantive indicator variables are compared under different constraints to determine (*a*) whether the latent MV has (un)equal effects on the substantive variables and (*b*) whether the relationships between the substantive variables are biased due to the latent MV (for a detailed description, see Williams et al. 2010).

The CFA-based MV technique is clearly more sophisticated than either the correlation-based or regression-based technique because these latter techniques (*a*) account for method variance only at the construct level, not at the level of the indicators; (*b*) do not control for measurement error; and (*c*) assume that CMV has equal effects on all observed variables, which is not supported by the available evidence (Baumgartner & Steenkamp 2001, Cote & Buckley 1987, Williams et al. 2010). In addition, the correlation-based MV technique assumes that CMV can only inflate (and not deflate) correlations among the substantive variables and that the regression-based technique applies only to single-equation models. Nevertheless, all three of the MV techniques have important limitations. First, it is unclear what specific sources of CMV the MV techniques are controlling (Simmering et al. 2015). Second, it is doubtful that any given MV will control for all potential sources of CMB (Williams & McGonagle 2016). More importantly, it is unlikely that CMB that is attributable to relationship-specific sources of CMB (e.g., implicit theories and consistency motifs) is controlled using this technique, because the MV is supposed to be unrelated to the substantive variables included in a study. This limitation is important because implicit theories have been shown to influence a variety of substantive relationships in the field (Eden & Leviatan 1975, Lord et al. 1978, Smither et al. 1989), and there is considerable evidence that people try to maintain consistency in their responses to events/survey items. Finally, several researchers (Baumgartner & Weijters 2021, Steenkamp & Maydeu-Olivares 2021) have noted that the practice of selecting MVs often does not satisfy the criteria identified by Richardson et al. (2009) and does not include the emphasis on linking to measurement theory advocated by Williams et al. (2010).

Directly measured latent variable technique. The DMLV technique improves on the CFA marker technique by using measures that are presumed to directly represent processes that generate CMV. Williams & Anderson (1994) and Williams et al. (1996) originally used the term measured method effect variables in this context, and they delineated how the technique should be implemented. More recently, Williams & McGonagle (2016) used the term measured cause to refer to DMLVs. To apply the DMLV technique, researchers identify the specific source of CMB that they believe might be present in a study, directly measure it, and control for it in the analyses. As such, the form of the CFA model for the DMLV technique is the same as for the CFA marker model shown in **Figure 7*d***. The only difference is in the nature of the variables used in the attempt to detect and control for CMV. The most common variables included by researchers when using this technique include social desirability (Barrick & Mount 1996), positive or negative affect (Schaubroeck et al. 1992, Williams & Anderson 1994, Williams et al. 1996), impression management (Brady et al. 2017), and response styles (Baumgartner & Steenkamp 2001, Weijters et al. 2008). Once the specific form of CMB that is expected to influence the substantive relationships has been identified and measured, (*a*) the method effect indicators included in the study are allowed to load on their theoretical constructs and on a latent method factor that has its own measurement component, and (*b*) the significance of the factor correlations or structural parameter estimates is examined both with and without the latent method factor in the model. If the parameter estimates remain significant with the method factor included, the researcher can conclude that the structural relationships are supported.

Among the advantages of the DMLV approach are that it identifies the specific source of CMB that is expected to affect the substantive relationships, it controls for the effects of CMB at the item level rather than the construct level, it controls for measurement error, and it does not constrain the loadings on method factors to be equal. However, this technique is not without limitations. First, it requires the researcher to identify the specific source of the CMB and to be able to measure it. Second, it implicitly assumes that the source of CMB identified and included in the analysis is the most important one in that research setting, which may not be true. Third, trying to control for multiple sources of bias using this technique only complicates things because it means having to identify these sources and obtain adequate measures of them. Finally, like the other statistical techniques discussed above, it is unlikely that this technique can be used to control for some of the most potent causes of CMB (e.g., implicit theories and consistency motifs); however, for exceptions, see the attempts by Phillips & Lord (1986) and Rush et al. (1977) to measure and control for implicit leadership theories. With that said, we believe that the DMLV technique is valuable when the potential sources of CMB can be identified and validly measured.

Summary of Statistical Techniques

None of the most-used statistical techniques (HSF, the UMLV technique, and the variations of the MV technique) provide a satisfactory solution to controlling for CMB. In addition to being limited by the fact that they do not identify the specific source of CMB that may be causing bias in the data, they all have other inadequacies. HSF is based on an arbitrary criterion; is likely to confound substantive and method variance; and, most importantly, does not actually control CMB. Thus, we do not recommend the use of this technique in any situation. The UMLV technique is also likely to confound substantive and method variance, and research examining this technique (Chin et al. 2012, Richardson et al. 2009, Spector et al. 2019) has raised questions about its efficacy. Finally, although there are differences in the ways the MV technique has been implemented, many empirical tests using this remedy do not satisfy the criteria for choosing an effective MV, so it is doubtful that any given MV will control for all potential sources of CMB, and it is unlikely that this technique can control for some of the most potent sources of method variance (e.g., implicit theories, consistency motifs). The limitations of these statistical techniques are particularly important because recent studies concluding that CMB is unlikely to present a problem (Bozionelos & Simmering 2022, Fuller et al. 2016) relied heavily on evidence provided by studies using these techniques. We think that these conclusions are unfounded because they rely on techniques that are limited by the scope of CMB that they control and are ineffective at controlling these biases. In contrast, we feel the DMLV approach can be of great value, especially for research methodologists trying to understand specific sources of CMV and CMB. However, as is true of all studies using specific measures for latent variables representing method effects, conclusions are limited to the specific measures incorporated in these models. We cannot emphasize enough that the control of a few sources of method effects does not mean that the influence of others has been accounted for. Therefore, we recommend that researchers focus on the general and specific procedural remedies for controlling CMB in the designs of their studies, and use statistical techniques only when they can identify a specific biasing factor that can be measured.

FUTURE RESEARCH DIRECTIONS

Our review of the literature suggests that we have increased our understanding of the causes and consequences of CMB and its remedies. In this section, however, we discuss what we consider to be some of the more interesting avenues for future research (for a summary, see **Table 5**).

Table 5 Summary of future research directions

Future research direction	Example research questions	Example (or possible) studies
Experiments designed to examine the effects of method factors	How does item proximity affect CMV/CMB? What are the influences of item characteristics, item contextual factors, rater characteristics, and measurement context on CMV/CMB? What is the impact of transient mood on measures of work-related affect?	Wilson et al. (2021) used an experimental design to compare five item-ordering approaches: (*a*) individually randomized items (randomized), (*b*) static items grouped by construct (grouped), (*c*) static intermixed items (intermixed), (*d*) individually randomized grouped-by-construct blocks containing static items (random blocks), and (*e*) static grouped-by-construct blocks containing individually randomized items (static blocks).
Research on multidimensional constructs	What are the unique challenges related to CMV for multidimensional constructs? Are measures of different dimensions of a multidimensional construct contaminated by common CMV processes, and what is the impact of efforts to establish discriminant validity? What are the most effective ways to detect and control CMB when studying multidimensional constructs?	Johnson et al. (2011) examined the extent to which CMB inflates the relationship between core self-evaluation (a higher-order construct) and job satisfaction using both statistical and procedural remedies.
Research on multilevel models	What are the challenges related to CMV for multilevel models? Are measures of group-level variables obtained via individual self-reports affected by different sources of CMV compared with individual-level variables? What procedures are most useful for detecting and controlling CMV in multilevel models?	Lai et al. (2013) found that although CMV is unlikely to generate significant cross-level interaction effects, it may lead to the identification of false significant cross-level main effects when no true main effect exists.
Experience sampling studies	What sources of CMV are especially worth considering when conducting experience sampling studies? What sources of CMV are the most problematic with ESM, given that the data are separated in time but the same measures are typically used repeatedly? What procedural (or statistical) remedies are useful for reducing the influences of CMV in experience sampling studies?	Gabriel et al. (2019, question 8) noted that the effects of mood states or emotions may be controlled by measuring them and partialling out their effects or by analyzing the lagged relationships between the predictor and criterion variables.
Cross-cultural studies	How does culture influence the sources and strengths of CMV? How do cultural differences in causes of CMV affect tests of measurement equivalence needed in cross-cultural studies? What procedures are useful for reducing the potential impacts of CMV in certain cultures?	Cultural differences have been observed in socially desirable responding (Lalwani et al. 2009), implicit theories (Church et al. 2012), and positive and negative affect (Bagozzi et al. 1999).

Abbreviations: CMB, common method bias; CMV, common method variance; ESM, experience sampling method.

Experiments Designed to Examine the Effects of Method Factors

Although good examples of experimental studies on CMB exist (e.g., Harrison & McLaughlin 1993; Weijters et al. 2009, 2014; A.L. Rubenstein, personal communication, Sept. 6, 2022), a recent study by Wilson et al. (2021) demonstrates the complexity of the causes of these biases, as well as some of the trade-offs that researchers must consider when trying to control for their effects. Wilson et al. examined the effects of item ordering on the reproducibility of online survey studies. Specifically, these authors compared five item-ordering approaches: (*a*) individually randomized items (randomized), (*b*) static items grouped by construct (grouped), (*c*) static intermixed items (intermixed), (*d*) individually randomized grouped-by-construct blocks containing static items (random blocks), and (*e*) static grouped-by-construct blocks containing individually randomized items (static blocks). They found that (*a*) the average reliability of the measures included in their study was significantly higher for colocated construct items (grouped, static block, or random block item ordering) than for randomized or intermixed items, (*b*) the construct validity of the measures of the focal constructs (based on the number of violations of convergent and discriminant validity of the measures) was higher for static block and random block item ordering than for the three other types of item ordering, and (*c*) respondents reported less fatigue and frustration in the grouped versus intermixed conditions as well as lower fatigue in the grouped versus randomized conditions. Although the authors noted that increased fatigue and frustration are natural outcomes of increased cognitive effort (which would be beneficial in reducing the tendency to satisfice), their other findings suggest that block designs are better item-ordering options for researchers to use.

Researchers should exercise caution here. Although grouping items or using block designs may decrease respondent fatigue and frustration and improve the psychometric properties of a measuring instrument, the associated inflation of estimates of construct reliability and validity is problematic. Therefore, although grouping items of a scale may make sense after a measure has been validated, we agree with Wilson et al. (2021) that intermixing or randomizing items during the instrument development and validation phases is likely to produce a more robust measure for future research. Thus, as in any research, it is important to consider the trade-offs that one must make while also controlling for CMB.

In addition to manipulating the relative proximity of item wording, researchers could manipulate (*a*) item characteristics, such as the nature and number of item response options provided to participants (Lambert et al. 2022); (*b*) item context factors, such as grammatical redundancy within and across measures of constructs (Cortina et al. 2020) as well as item priming and embeddedness; (*c*) rater characteristics, such as transient mood state; and (*d*) measurement context, in the form of temporal, physical, or psychological separation. Researchers could also use experiments to provide evidence for the measurement-specific approach to understanding method variance and biases (Spector et al. 2019, 2022) by manipulating the above factors in administrations of different measures of the same, and different, construct(s).

Research on Multidimensional Constructs

We also encourage additional research on the effects that CMB has on higher-order multidimensional constructs, given their increased popularity. Johnson et al. (2011) noted that multidimensional constructs present unique challenges because researchers must examine the extent to which method factors affect the relationships among the lower-level dimensions of the higher-order construct as well as the relationships between the higher-order focal construct and other constructs in its nomological network; moreover, each dimension added to a multidimensional construct may introduce new sources of CMB. Johnson et al. examined the extent to which

CMB inflates the relationships between CSE (a higher-order construct) and job satisfaction using both statistical and procedural remedies. In terms of statistical remedies, these authors concluded that controlling for a measured (e.g., social desirability) or an unmeasured latent method factor appeared to be more effective at reducing CMB than controlling for an MV. In terms of procedural remedies, they found that temporal separation, particularly temporal separation of each of the lower-level dimensions of the higher-order CSE construct, was more effective than using a combination of different response formats and filler scales between the subdimensions of the higher-order CSE construct. However, it is unclear how many of these findings were due to sampling differences, or whether they would generalize to other higher-order constructs. Therefore, we encourage more research focusing on the effects of CMB on other multidimensional constructs. Although several such constructs come to mind (overall job satisfaction, the Big Five personality traits, challenge and hindrance stressors), given the important role that overall job performance plays in the fields of management and applied psychology, and the fact that it is often conceptualized as consisting of both positive (task performance and OCB) and negative (CWB) subdimensions (Rotundo & Sackett 2002) that may cause several different types of CMB (e.g., trait and item social desirability/undesirability, implicit theories, leniency biases), it may prove to be a particularly interesting construct to explore in future research.

Research on Multilevel Models

Several articles (Mathieu & Chen 2011, Mathieu et al. 2012, Ostroff et al. 2002) have chronicled the increased application of multilevel theories, designs, and analyses in the organizational sciences. As noted by Lai et al. (2013), one of the main concerns in multilevel models is the potential confounding effects of CMV on cross-level interactions—in other words, the moderating effects of higher-level variables on the relationships between lower-level variables. Preliminary research by these authors using a simulation study suggests that, in the absence of true effects, CMV is unlikely to generate significant cross-level interactive effects or bias these parameter estimates. Indeed, similar to the conclusion obtained by Siemsen et al. (2010) from their examination of individual-level interactive effects, Lai et al. found that CMV tends to suppress the identification of true cross-level interactions. However, these authors found that CMV may lead to the identification of false significant cross-level main effects and overestimation of the regression coefficient when no true effects exist. Given their findings, Lai et al. encourage systematic research that aims to identify the mechanisms that explain the underestimation of true cross-level interaction effects and the effects of CMV on cross-level main effects. We agree. However, we also feel that additional primary studies, including (when possible) experiments designed to examine the effects of CMB on multilevel relationships, should prove worthwhile.

Experience Sampling Studies

Gabriel et al. (2019) have reported on the rapid growth in experience sampling method (ESM) studies in the fields of organizational behavior and applied psychology. This growth reflects increased interest in intraindividual (i.e., within-person) as opposed to interindividual (i.e., between-person) phenomena during the past few decades. As noted by Gabriel et al. (2019, question 8), ESM studies often raise concerns about the potential effects of CMV because most of these studies gather the predictor and criterion variables over multiple administrations from the same rating source. These authors indicate that the most widely used remedy for addressing CMV in these studies is person-mean (or group-mean) centering, because this procedure controls for virtually all forms of between-person differences (e.g., demographics, personality, response tendencies, social desirability). Unfortunately, however, person-mean centering does not control for potential biases related to transient mood states, emotions, or affect, nor does it control for the possible effects of item characteristics or item contexts. Gabriel et al. (2019, question 8) note that the effects of mood states or emotions may be controlled by measuring them and partialling out their effects or by analyzing the lagged relationships between the predictor and criterion variables. However, these remedies are not without limitations (for a more complete discussion of the issues, see Gabriel et al. 2019, question 8). In addition, these authors note that little research has focused on the effects that item characteristics have on ESM study results. Thus, additional research is needed to determine the effects that these factors have on the growing number of studies using this methodology.

Cross-Cultural Studies

Given the growing number of cross-cultural comparison studies that have been conducted during the past 15 years (Broesch et al. 2020), and the fact that most of the studies reported in management and applied psychology adapt questionnaires developed from English-speaking countries (Harzing 2006, Harzing et al. 2012), perhaps it is not surprising that there is an increased interest in the effects of cultural differences on CMB. However, it is surprising that most of the research in this area has focused on how response styles differ across cultures (e.g., Baumgartner & Steenkamp 2001, Benitez et al. 2016, Harzing et al. 2012). Although we believe that such research is important, we also believe that it is too restrictive in its scope and that there is a need for additional research on the effects that cultural differences have on CMB. For example, the fact that cultural differences have been found in socially desirable responding (Lalwani et al. 2009), implicit theories (Church et al. 2012), and positive and negative affect (Bagozzi et al. 1999) suggests that additional research should focus on the effects that culture has on CMB associated with rating sources.

CONCLUSION

It has been 30 years since Schmitt (1994) noted that in order to develop a better understanding of the effects of CMB, researchers in the field of applied psychology need to make a stronger commitment to (*a*) identify the sources of CMB, (*b*) clarify how these sources affect the nature of substantively interesting relationships, and (*c*) explain how CMB could be measured and/or controlled. We believe that the field has made considerable strides in addressing these issues (see the Summary Points). For example, during the past few decades researchers have devoted significant effort to identifying and categorizing the various sources of CMB (e.g., Baumgartner et al. 2021; MacKenzie & Podsakoff 2012; Podsakoff et al. 2003, 2012; Weijters et al. 2009, 2010). Relatedly, researchers have made a concerted attempt to uncover the mechanisms through which CMB has its effects (MacKenzie & Podsakoff 2012, Podsakoff et al. 2012, Yao & Xu 2021) and the stages in which CMB is likely to enter the survey response process (Podsakoff et al. 2003). We also have a much better understanding of the effects that CMB can have on measures of construct validity and reliability (Baumgartner & Steenkamp 2001, Cote & Buckley 1987, Williams et al. 2010) and on the nature of the relationships between constructs (Baumgartner et al. 2021; Podsakoff et al. 2003, 2012; Williams et al. 2010). Moreover, various procedural and statistical remedies to control CMB (Jakobsen & Jensen 2015, MacKenzie & Podsakoff 2012) have been identified and implemented, and there is growing evidence about which of these techniques work and which ones do not (Baumgartner & Weijters 2021, Baumgartner et al. 2021, Hulland et al. 2018). This knowledge should be gratifying to those researchers interested in understanding and trying to control biases in their research. Thus, even though we have not resolved all the issues related to this topic, we hope that we have provided a worthwhile summary of what we know about the sources of CMB, its effects, the remedies that can be used to address it, and some directions for future research.

SUMMARY POINTS

1. Despite recognition that method biases can have several harmful effects, the causes and consequences of common method bias (CMB), and remedies for dealing with it, are still not well understood.
2. Method factors are bad in that they can have harmful effects on estimates of construct reliability and validity and on the empirical relationships between measures of different constructs. CMB is more likely to be problematic when the rater lacks the ability or the motivation to provide accurate ratings, the task (survey questionnaire) is difficult, and/or raters are given the opportunity to exert low effort (i.e., satisfice).
3. CMB is complex for two reasons: It can result from several different sources, including rater characteristics, item characteristics, item context effects, and the measurement context, and it can inflate, deflate, or have no effect on the observed relationships between focal constructs.
4. CMB appears to be widespread, as several disciplines, including applied psychology, organizational behavior, marketing, operations management, and management information systems, have reported that between 31% and 98% of published studies use designs that are susceptible to it.
5. CMB is not easy to fix because many of the most commonly used statistical techniques have substantial limitations, some remedies address one or only a few sources of potential CMB but do not address other sources of CMB, and some techniques may mitigate one or more sources of CMB while simultaneously magnifying the effects of other sources of CMB.
6. The most widely used statistical remedy (i.e., Harman's single-factor technique) is limited and does not control for CMB, and we recommend strongly that its use be discontinued.
7. When possible, researchers should use procedural remedies to control CMB. By obtaining measures of the predictor and criterion variables from different sources and including a temporal separation between them, researchers can effectively negate several of the major causes of CMB.
8. Avenues for future research that should prove beneficial include experiments designed to test the effects of different sources and remedies of CMB, remedies for CMB in measures of multidimensional constructs, remedies for CMB in multilevel models and designs using experience sampling methods, and comparisons of CMB effects across cultures.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

P.M.P. and N.P.P. acknowledge the influence of Scott MacKenzie on this article. Although Scott did not participate in its writing, his intellectual contributions will be evident to those readers who are familiar with his previous collaborations with the first two authors on the topic of common method bias.

LITERATURE CITED

Ackerman C, Warren MA, Donaldson SI. 2018. Scaling the heights of positive psychology: systematic review of measurement scales. *Int. J. Wellbeing* 8:1–21

Aguirre-Urreta MI, Hu J. 2019. Detecting common method bias: performance of the Harman's single-factor test. *Database Adv. Inf. Syst.* 50:45–70

Anseel F, Lievens F, Schollaert E, Choragwicka B. 2010. Response rates in organizational science, 1995–2008: a meta-analytic review and guidelines for survey researchers. *J. Bus. Psychol.* 25:335–49

Antonakis J, Bendahan S, Jacquart P, Lalive R. 2010. On making causal claims: a review and recommendations. *Leadersh. Q.* 21:1086–120

Astra RL, Singg S. 2000. The role of self-esteem in affiliation. *J. Psychol.* 134:15–22

Bachrach DG, Bendoly E, Podsakoff PM. 2001. Attributions of the "causes" of group performance as an alternative explanation of the relationship between organizational citizenship behavior and organizational performance. *J. Appl. Psychol.* 86:1285–93

Bagozzi RP. 1984. A prospectus for theory construction in marketing. *J. Mark.* 48:11–29

Bagozzi RP. 2011. Measurement and meaning in information systems and organizational research: methodological and philosophical foundations. *MIS Q.* 35:261–91

Bagozzi RP, Li YJ, Phillips LW. 1991. Assessing construct-validity in organizational research. *Adm. Sci. Q.* 36:421–58

Bagozzi RP, Wong N, Li YJ. 1999. The role of culture and gender in the relationship between positive and negative affect. *Cogn. Emot.* 13:641–72

Bagozzi RP, Yi YJ. 1990. Assessing method variance in multitrait multimethod matrices: the case of self-reported affect and perceptions at work. *J. Appl. Psychol.* 75:547–60

Barrick MR, Mount MK. 1996. Effects of impression management and self-deception on the predictive validity of personality constructs. *J. Appl. Psychol.* 81:261–72

Baumgartner H, Steenkamp J. 2001. Response styles in marketing research: a cross-national investigation. *J. Mark. Res.* 38:143–56

Baumgartner H, Weijters B. 2021. Dealing with common method variance in international marketing research. *J. Int. Mark.* 29:7–22

Baumgartner H, Weijters B, Pieters R. 2021. The biasing effect of common method variance: some clarifications. *J. Acad. Mark. Sci.* 49:221–35

Beal DJ, Weiss HM, Barros E, MacDermid SM. 2005. An episodic process model of affective influences on performance. *J. Appl. Psychol.* 90:1054–68

Benitez I, He J, Van de Vijver FJR, Padilla J-L. 2016. Linking extreme response style to response processes: a cross-cultural mixed methods approach. *Int. J. Psychol.* 51:464–73

Bliese PD. 2000. Within-group agreement, non-independence, and reliability: implications for data aggregation and analysis. In *Multilevel Theory, Research, and Methods Inorganizations: Foundations, Extensions and New Directions*, ed. KJ Klein, SW Kozlowski, pp. 349–81. San Francisco, CA: Jossey-Bass

Bodner TE. 2006. Designs, participants, and measurement methods in psychological research. *Can. Psychol.* 47:263–72

Boehm SA, Kunze F, Bruch H. 2014. Spotlight on age-diversity climate: the impact of age-inclusive HR practices on firm-level outcomes. *Pers. Psychol.* 67:667–704

Bollen KA. 1989. *Structural Equations with Latent Variables*. New York: Wiley

Bozionelos N, Simmering MJ. 2022. Methodological threat or myth? Evaluating the current state of evidence on common method variance in human resource management research. *Hum. Resour. Manag. J.* 32:194–215

Brady DL, Brown DJ, Liang LH. 2017. Moving beyond assumptions of deviance: the reconceptualization and measurement of workplace gossip. *J. Appl. Psychol.* 102:1–25

Brannick MT, Chan D, Conway JM, Lance CE, Spector PE. 2010. What is method variance and how can we cope with it? A panel discussion. *Organ. Res. Methods* 13:407–20

Broesch T, Crittenden AN, Beheim BA, Blackwell AD, Bunce JA, et al. 2020. Navigating cross-cultural research: methodological and ethical considerations. *Proc. R. Soc. B* 287:20201245

Brown TA. 2015. *Confirmatory Factor Analysis for Applied Research*. New York: Guilford. 2nd ed.

Buckley MR, Cote JA, Comstock SM. 1990. Measurement errors in behavioral sciences: the case of personality/attitude research. *Educ. Psychol. Meas.* 50:447–74

Burton-Jones A. 2009. Minimizing method bias through programmatic research. *MIS Q.* 33:445–71

Campbell DT, Fiske DW. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56:81–105

Campbell DT, O'Connell EJ. 1967. Methods factors in multitrait-multimethod matrices: multiplicative rather than additive. *Multivar. Behav. Res.* 2:409–26

Chaffin D, Heidl R, Hollenbeck JR, Howe M, Yu A, et al. 2017. The promise and perils of wearable sensors in organizational research. *Organ. Res. Methods* 20:3–31

Chan D. 1998. Functional relations among constructs in the same content domain at different levels of analysis: a typology of composition models. *J. Appl. Psychol.* 83:234–46

Chang SJ, van Witteloostuijn A, Eden L. 2010. From the editors: common method variance in international business research. *J. Int. Bus. Stud.* 41:178–84

Chen PY, Dai T, Spector PE, Jex SM. 1997. Relation between negative affectivity and positive affectivity: effects of judged desirability of scale items and respondents' social desirability. *J. Personal. Assess.* 69:183–98

Cheng KHC, Hui CH, Cascio WF. 2017. Leniency bias in performance ratings: the Big-Five correlates. *Front. Psychol.* 8:521

Chin WW, Thatcher JB, Wright R. 2012. Assessing common method bias: problems with the ULMC technique. *MIS Q.* 36:1003–19

Church AT, Willmore SL, Anderson AT, Ochiai M, Porter N, et al. 2012. Cultural differences in implicit theories and self-perceptions of traitedness: replication and extension with alternative measurement formats and cultural dimensions. *J. Cross-Cult. Psychol.* 43:1268–96

Chyung AY, Barkin JR, Shamsy JA. 2018. Evidence-based survey design: the use of negatively worded items in surveys. *Perform. Improv.* 57:16–25

Colquitt JA, LePine JA, Zapata CP, Wild RE. 2011. Trust in typical and high-reliability contexts: building and reacting to trust among firefighters. *Acad. Manag. J.* 54:999–1015

Connolly JJ, Viswesvaran C. 2000. The role of affectivity in job satisfaction: a meta-analysis. *Personal. Individ. Differ.* 29:265–81

Cooper B, Eva N, Fazlelahi FZ, Newman A, Lee A, Obschonka M. 2020. Addressing common method variance and endogeneity in vocational behavior research: a review of the literature and suggestions for future research. *J. Vocat. Behav.* 121:103472

Cortina JM, Sheng Z, Keener KK, Keeler KR, Grubb LH, et al. 2020. From alpha to omega and beyond! A look at the past, present, and (possible) future of psychometric soundness in the *Journal of Applied Psychology*. *J. Appl. Psychol.* 105:1351–81

Cote JA, Buckley MR. 1987. Estimating trait, method, and error variance: generalizing across 70 construct-validation studies. *J. Mark. Res.* 24:315–18

Cote JA, Buckley MR. 1988. Measurement error and theory testing in consumer research: an illustration of the importance of construct-validation. *J. Consum. Res.* 14:579–82

Craighead CW, Ketchen DJ, Dunn KS, Hult GTM. 2011. Addressing common method variance: guidelines for survey research on information technology, operations, and supply chain management. *IEEE Trans. Eng. Manag.* 58:578–88

Cram WA, D'Arcy J, Proudfoot JG. 2019. Seeing the forest and the trees: a meta-analysis of the antecedents to information security policy compliance. *MIS Q.* 43:525–54

Crampton SM, Wagner JA. 1994. Percept-percept inflation in microorganizational research: an investigation of prevalence and effect. *J. Appl. Psychol.* 79:67–76

Cruz KS. 2022. Are you asking the correct person (hint: Oftentimes you are not!)? Stop worrying about unfounded common method bias arguments and start using my guide to make better decisions of when to use self- and other-reports. *Group Organ. Manag.* 47:920–27

Cui TX, Kam CCS, Cheng EH, Ho MY. 2022. Distinguishing between trait desirability and item desirability in predicting item scores: Is informant evaluation of personality free from social desirability? *Personal. Individ. Differ.* 196:111708

Dalal RS. 2005. A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *J. Appl. Psychol.* 90:1241–55

Deng H, Lam CK, Guan Y, Wang M. 2021. My fault or yours? Leaders' dual reactions to abusive supervision via rumination depend on their independent self-construal. *Pers. Psychol.* 74:773–98

Detert JR, Edmondson AC. 2011. Implicit voice theories: taken-for-granted rules of self-censorship at work. *Acad. Manag. J.* 54:461–88

Doty DH, Glick WH. 1998. Common methods bias: Does common methods variance really bias results? *Organ. Res. Methods* 1:374–406

Eden D, Leviatan U. 1975. Implicit leadership theory as a determinant of factor structure underlying supervisory behavior scales. *J. Appl. Psychol.* 60:736–41

Edwards AL. 1953. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. Appl. Psychol.* 37:90–93

Edwards AL. 1957. *The Social Desirability Variable in Personality Assessment and Research*. New York: Dryden

Evans MG. 1985. A Monte-Carlo study of the effects of correlated method variance in moderated multiple-regression analysis. *Organ. Behav. Hum. Decis. Process.* 36:305–23

Fornell C, Larcker DF. 1981. Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18:39–50

Fuller CM, Simmering MJ, Atinc G, Atinc Y, Babin BJ. 2016. Common methods variance detection in business research. *J. Bus. Res.* 69:3192–98

Gabriel AS, Podsakoff NP, Beal DJ, Scott BA, Sonnentag S, et al. 2019. Experience sampling methods: a discussion of critical trends and considerations for scholarly advancement. *Organ. Res. Methods* 22:969–1006

Ganster DC, Rosen CC. 2013. Work stress and employee health: a multidisciplinary view. *J. Manag.* 39:1085–122

Gielnik MM, Zacher H, Wang M. 2018. Age in the entrepreneurial process: the role of future time perspective and prior entrepreneurial experience. *J. Appl. Psychol.* 103:1067–85

Grant AM, Hofmann DA. 2011. Outsourcing inspiration: the performance effects of ideological messages from leaders and beneficiaries. *Organ. Behav. Hum. Decis. Process.* 116:173–87

Greenberger E, Chen C, Dmitieva J, Farruggia SP. 2003. Item-wording and the dimensionality of the Rosenberg self-esteem scale: Do they matter? *Personal. Individ. Differ.* 35:1241–54

Greenwald AG, Farnham SD. 2000. Using the Implicit Association Test to measure self-esteem and self-concept. *J. Personal. Soc. Psychol.* 79:1022–38

Harris MM, Bladen A. 1994. Wording effects in the measurement of role-conflict and role ambiguity: a multitrait-multimethod analysis. *J. Manag.* 20:887–901

Harrison DA, McLaughlin ME. 1993. Cognitive-processes in self-report responses: tests of item context effects in work attitude measures. *J. Appl. Psychol.* 78:129–40

Harrison DA, McLaughlin ME, Coalter TM. 1996. Context, cognition, and common method variance: psychometric and verbal protocol evidence. *Organ. Behav. Hum. Decis. Process.* 68:246–61

Harvey RJ, Billings RS, Nilan KJ. 1985. Confirmatory factor-analysis of the job diagnostic survey: good news and bad news. *J. Appl. Psychol.* 70:461–68

Harzing AW. 2006. Response styles in cross-national survey research: a 26-country study. *Int. J. Cross-Cult. Manag.* 6:243–66

Harzing AW, Brown M, Koster K, Zhao S. 2012. Response style differences in cross-cultural research: dispositional and situational determinants. *Manag. Int. Rev.* 52:341–63

Heberlein TA, Baumgartner R. 1978. Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. *Am. Sociol. Rev.* 43:447–62

Heggestad ED, Scheaf DJ, Banks GC, Hausfeld MM, Tonidandel S, Williams EB. 2019. Scale adaptation in organizational science research: a review and best-practice recommendations. *J. Manag.* 45:2596–27

Henderson DJ, Liden RC, Gibkowski BC, Chaudry A. 2009. LMX differentiation: a multilevel review and examination of its antecedents and outcomes. *Leadersh. Q.* 20:517–34

Hopp WJ, Iravani SMR, Liu F. 2009. Managing white-collar work: an operations-oriented survey. *Prod. Oper. Manag.* 18:1–32

Huang C. 2013. Relation between self-esteem and socially desirable responding and the role of socially desirable responding in the relation between self-esteem and performance. *Eur. J. Psychol. Educ.* 28:663–83

Hulland J, Baumgartner H, Smith KM. 2018. Marketing survey research best practices: evidence and recommendations from a review of *JAMS* articles. *J. Acad. Mark. Sci.* 46:92–108

Ibrahim AM. 2001. Differential responding to positive and negative items: the case of a negative item in a questionnaire for course and faculty evaluation. *Psychol. Rep.* 88:497–500

Jakobsen M, Jensen R. 2015. Common method bias in public management studies. *Int. Public Manag. J.* 18:3–30

Janiszewski C, Wyer RS Jr. 2014. Content and process priming: a review. *J. Consum. Psychol.* 24:96–118

Johnson RE, Rosen CC, Djurdjevic E. 2011. Assessing the impact of common method variance on higher order multidimensional constructs. *J. Appl. Psychol.* 96:744–61

Jordan PJ, Troth AC. 2020. Common method bias in applied settings: the dilemma of researching in organizations. *Aust. J. Manag.* 45:3–14

Joseph DL, Chan MY, Heintzlman SJ, Tay L, Diener E, Scotney VS. 2020. The manipulation of affect: a meta-analysis of affect induction procedures. *Psychol. Bull.* 146:355–75

Judd C, Drake R, Downing J, Krosnick JA. 1991. Some dynamic properties of attitude structures: context-induced response facilitation and polarization. *J. Personal. Soc. Psychol.* 60:193–202

Kaltsonoudi K, Tsigilis N, Karteroliotis K. 2021. Critical review of the literature and current tendencies of the common method variance in sport management research. *Meas. Phys. Educ. Exerc. Sci.* 26:103–15

Kam CCS. 2013. Probing item social desirability by correlating personality items with Balanced Inventory of Desirable Responding (BIDR): a validity examination. *Personal. Individ. Differ.* 54:513–18

Kam CCS, Meyer JP. 2015. Implications of item keying and item valence for the investigation of construct dimensionality. *Multivar. Behav. Res.* 50:457–69

Kaufmann L, Saw AA. 2014. Using a multiple-informant approach in SCM research. *Int. J. Phys. Distrib. Logist. Manag.* 44:511–27

Kenny DA, Kashy DA. 1992. Analysis of multitrait-multimethod matrix by confirmatory factor analysis. *Psychol. Bull.* 112:165–72

Ketokivi MA, Schroeder RG. 2004. Perceptual measures of performance: fact or fiction? *J. Oper. Manag.* 22:247–64

Kock F, Berbekova A, Assaf AG. 2021. Understanding and managing the threat of common method bias: detection, prevention, and control. *Tour. Manag.* 86:104330

Krosnick JA. 1991. Response strategies for coping with the demands of attitude measures in surveys. *Appl. Cogn. Psychol.* 5:213–36

Krosnick JA. 1999. Survey research. *Annu. Rev. Psychol.* 50:537–67

Lai X, Li FL, Leung K. 2013. A Monte Carlo study of the effects of common method variance on significance testing and parameter bias in Hierarchical Linear Modeling. *Organ. Res. Methods* 16:243–46

Lalwani AK, Shrum LJ, Chiu C-Y. 2009. Motivated response styles: the role of cultural values, regulatory focus, and self-consciousness in socially desirable responding. *J. Personal. Soc. Psychol.* 96:870–82

Lambert LS, Gray T, Davis A, Erdman M, McDermott R. 2022. *An overlooked aspect of measurement: Does the content of verbal anchors matter?* Paper presented at Society for Industrial and Organizational Psychology Annual Conference, Seattle, WA, April 27–30

Lance CE, Dawson B, Birkelbach D, Hoffman BJ. 2010. Method effects, measurement error, and substantive conclusions. *Organ. Res. Methods* 13:435–55

Le H, Schmidt FL, Putka DJ. 2009. The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organ. Res. Methods* 12:165–200

Lelkes Y, Krosnick JA, Max D, Judd C, Park B. 2012. Complete anonymity compromises the accuracy of self-reports. *J. Exp. Soc. Psychol.* 48:1291–99

Lench HC, Flores SA, Bench SW. 2011. Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. *Psychol. Bull.* 137:834–55

Liang H, Saraf N, Hu Q, Xue Y. 2007. Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. *MIS Q.* 31:59–87

Lim S, Ilies R, Koopman J, Christoforou P, Arvey RD. 2018. Emotional mechanisms linking incivility at work to aggression and withdrawal at home: an experience-sampling study. *J. Manag.* 44:2888–908

Lindell MK, Whitney DJ. 2001. Accounting for common method variance in cross-sectional research designs. *J. Appl. Psychol.* 86:114–21
Little TD, Lindenberger U, Nesselrode JR. 1999. On selecting indicators for multivariate measurement and modeling of latent variables: when 'good' indicators are bad and 'bad' indicators are good. *Psychol. Methods* 4:192–211
Loi TI, Kuhn KM, Sahaym A, Butterfield KD, Tripp TM. 2020. From helping hands to harmful acts: when and how employee volunteering promotes workplace deviance. *J. Appl. Psychol.* 105:944–58
Lord RG, Binning JF, Rush MC, Thomas JC. 1978. The effect of performance cues and leader-behavior on questionnaire ratings of leadership behavior. *Organ. Behav. Hum. Perform.* 21:27–39
MacKenzie SB, Podsakoff PM. 2012. Common method bias in marketing: causes, mechanisms, and procedural remedies. *J. Retail.* 88:542–55
MacKenzie SB, Podsakoff PM, Podsakoff NP. 2011. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. *MIS Q.* 35:293–34
Magazine SL, Williams LJ, Williams ML. 1996. A confirmatory factor analysis examination of reverse coding effects in Meyer and Allen's affective and continuance commitment scales. *Educ. Psychol. Meas.* 56:241–50
Martin R, Thomas G, Legood A, Dello Russo S. 2018. Leader–member exchange (LMX) differentiation and work outcomes: conceptual clarification and critical review. *J. Organ. Behav.* 39:151–68
Mathews BP, Shepherd JL. 2002. Dimensionality of Cook and Wall's 1980 British organizational commitment questionnaire. *J. Occup. Organ. Psychol.* 75:369–75
Mathieu JE, Aguinis H, Culpepper SA, Chen G. 2012. Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *J. Appl. Psychol.* 97:951–66
Mathieu JE, Chen G. 2011. The etiology of the multilevel paradigm in management research. *J. Manag.* 37:610–41
McDermott MS, Sharma R. 2017. Evaluating the impact of method bias in health behaviour research: a meta-analytic examination of studies utilising the theories of reasoned action and planned behaviour. *Health Psychol. Rev.* 11:358–73
Meier KJ, O'Toole LJ. 2013. Subjective organizational performance and measurement error: common source bias and spurious relationships. *J. Public Adm. Res. Theory* 23:429–56
Messick S. 1991. Psychology and methodology of response styles. In *Improving the Inquiry in Social Science: A Volume in Honor of Lee J. Cronbach*, ed. RE Snow, DE Wiley, pp. 161–200. Hillsdale, NJ: Erlbaum
Min H, Park J, Kim HJ. 2016. Common method bias in hospitality research: a critical review of literature and an empirical study. *Int. J. Hosp. Manag.* 56:126–35
Mishra DP. 2000. An empirical assessment of measurement error in health-care survey research. *J. Bus. Res.* 48:193–205
Mitchell TR, James LR. 2001. Building better theory: time and the specification of when things happen. *Acad. Manag. Rev.* 26:530–47
Modem R, Lakshminarayanan S, Pillai R, Prabhu N. 2022. Twenty-five years of career growth literature: a review and research agenda. *Ind. Commer. Train.* 54:152–82
Montabon F, Daugherty PJ, Chen HZ. 2018. Setting standards for single respondent survey design. *J. Supply Chain Manag.* 54:35–41
Morgeson FP, Hofmann DA. 1999. The structure and function of collective constructs: implications for multilevel research and theory development. *Acad. Manag. Rev.* 24:249–65
Nederhof AJ. 1985. Methods of coping with social desirability bias: a review. *Eur. J. Soc. Psychol.* 15:263–80
Nifadkar S, Tsui AS, Ashforth BE. 2012. The way you make me feel and behave: supervisor-triggered newcomer affect and approach-avoidance behavior. *Acad. Manag. J.* 55:1146–68
Ones DS, Viswesvaran C, Reiss AD. 1996. Role of social desirability in personality testing for personnel selection: the red herring. *J. Appl. Psychol.* 81:660–79
Ostroff C, Kinicki AJ, Clark MA. 2002. Substantive and operational issues of response bias across levels of analysis: an example of climate-satisfaction relationships. *J. Appl. Psychol.* 87:355–68
Phillips JS, Lord RG. 1986. Notes on the practical and theoretical consequences of implicit theories for the future of leadership research. *J. Manag.* 12:31–41
Podsakoff NP, Maynes TD, Whiting SW, Podsakoff PM. 2015. One (rating) from many (observations): factors affecting the individual assessment of voice behavior in groups. *J. Appl. Psychol.* 100:1189–202

Podsakoff NP, Spoelma TM, Chawla N, Gabriel AS. 2019. What predicts within-person variance in applied psychology constructs? An empirical examination. *J. Appl. Psychol.* 101:727–54
Podsakoff NP, Whiting SW, Welsh DT, Mai KM. 2013. Surveying for “artifacts”: the susceptibility of the OCB–performance evaluation relationship to common rater, item, and measurement context effects. *J. Appl. Psychol.* 98:863–74
Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88:879–903
Podsakoff PM, MacKenzie SB, Podsakoff NP. 2012. Sources of method bias in social science research and recommendations on how to control it. *Annu. Rev. Psychol.* 63:539–69
Podsakoff PM, Organ DW. 1986. Self-reports in organizational research: problems and prospects. *J. Manag.* 12:531–44
Pujol-Cols L, Lazzaro-Salazar M. 2021. Ten years of research on psychosocial risks, health, and performance in Latin America: a comprehensive systematic review and research agenda. *J. Work Organ. Psychol.* 37:187–202
Rafferty AE, Griffin MA. 2004. Dimensions of transformational leadership: conceptual and empirical extensions. *Leadersh. Q.* 15:329–54
Rafferty AE, Griffin MA. 2006. Refining individualized consideration: distinguishing developmental leadership and supportive leadership. *J. Occup. Organ. Psychol.* 79:37–61
Richardson HA, Simmering MJ, Sturman MC. 2009. A tale of three perspectives: examining post hoc statistical techniques for detection and correction of common method variance. *Organ. Res. Methods* 12:762–800
Roth PL, BeVier CA. 1998. Response rates in HRM/OB survey research: norms and correlates, 1990–1994. *J. Manag.* 24:97–117
Rotundo M, Sackett PR. 2002. The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: a policy-capturing approach. *J. Appl. Psychol.* 87:66–80
Rush MC, Thomas JC, Lord RG. 1977. Implicit leadership theory: a potential threat to the internal validity of leader behavior questionnaires. *Organ. Behav. Hum. Perform.* 20:93–110
Sackett PR, Larson JR Jr. 1990. Research strategies and tactics in industrial and organizational psychology. In *Handbook of Industrial and Organizational Psychology*, ed. MD Dunnette, LM Hough, pp. 419–89. Palo Alto, CA: Consult. Psychol.
Schaubroeck J, Ganster DC, Fox ML. 1992. Dispositional affect and work-related stress. *J. Appl. Psychol.* 77:322–35
Schmitt N. 1994. Method bias: the importance of theory and measurement. *J. Organ. Behav.* 15:393–98
Schmitt N, Stults DM. 1986. Methodology review: analysis of multitrait-multimethod matrices. *Appl. Psychol. Meas.* 10:1–22
Schriesheim CA. 1981a. The effect of grouping and randomizing items on leniency response bias. *Educ. Psychol. Meas.* 41:401–11
Schriesheim CA. 1981b. Leniency effects on convergent and discriminant validity for grouped questionnaire items: a further investigation. *Educ. Psychol. Meas.* 41:1093–99
Schulte M, Ostroff C, Shmulyian S, Kinicki A. 2009. Organizational climate configurations: relationships to collective attitudes, customer satisfaction, and financial performance. *J. Appl. Psychol.* 94:618–34
Schwarz A, Rizzuto T, Carraher-Wolverton C, Roldan JL, Barrera-Barrera R. 2017. Examining the impact and detection of the “Urban Legend” of common method bias. *Database Adv. Inf. Syst.* 48:93–119
Scott BA, Lennard AC, Mitchell RL, Johnson RE. 2020. Emotions naturally and laboriously expressed: antecedents, consequences, and the role of valence. *Pers. Psychol.* 763:587–613
Sessions H, Nahrgang JD, Newton DW, Chamberlin M. 2020. I’m tired of listening: the effects of supervisor appraisals of group voice on supervisor emotional exhaustion and performance. *J. Appl. Psychol.* 105:619–36
Shipp AJ, Cole MS. 2015. Time in individual-level organizational studies: What is it, how is it used, and why isn’t it exploited more often? *Annu. Rev. Organ. Psychol. Organ. Behav.* 2:237–60
Shipp AJ, Jansen KJ. 2021. The “other” time: a review of the subjective experience of time in organizations. *Acad. Manag. Annu.* 15:299–334
Siemsen E, Roth A, Oliveira P. 2010. Common method bias in regression models with linear, quadratic, and interaction effects. *Organ. Res. Methods* 13:456–76

Simmering MJ, Fuller CM, Richardson HA, Ocal Y, Atinc GM. 2015. Marker variable choice, reporting, and interpretation in the detection of common method variance: a review and demonstration. *Organ. Res. Methods* 18:473–511

Smith CA, Organ DW, Near JP. 1983. Organizational citizenship behavior: its nature and antecedents. *J. Appl. Psychol.* 68:653–63

Smither JW, Collins H, Buda R. 1989. When ratee satisfaction influences performance evaluations: a case of illusory correlation. *J. Appl. Psychol.* 74:599–605

Spector PE. 1987. Method variance as an artifact in self-reported affect and perceptions at work: myth or significant problem? *J. Appl. Psychol.* 72:438–43

Spector PE. 2006. Method variance in organizational research: truth or urban legend? *Organ. Res. Methods* 9:221–32

Spector PE. 2019. Do not cross me: optimizing the use of cross-sectional designs. *J. Bus. Psychol.* 34:125–37

Spector PE, Bauer JA, Fox S. 2010. Measurement artifacts in the assessment of counterproductive work behavior and organizational citizenship behavior: Do we know what we think we know? *J. Appl. Psychol.* 95:781–90

Spector PE, Brannick MT. 2010. Common method issues: an introduction to the feature topic in *Organizational Research Methods*. *Organ. Res. Methods* 13:403–6

Spector PE, Gray CE, Rosen CC. 2022. Are biasing factors idiosyncratic to measures? A comparison of interpersonal conflict, organizational constraints, and workload. *J. Bus. Psychol.* 38:983–1002

Spector PE, Nixon AE. 2019. How often do I agree: an experimental test of item format method variance in stress measures. *Occup. Health Sci.* 3:125–43

Spector PE, Pindek S. 2016. The future of research methods in work and occupational health psychology. *Appl. Psychol.* 65:412–31

Spector PE, Rosen CC, Richardson HA, Williams LJ, Johnson RE. 2019. A new perspective on method variance: a measure-centric approach. *J. Manag.* 45:855–80

Staw BM. 1975. Attribution of the "causes" of performance: a general alternative interpretation of cross-sectional research on organizations. *Organ. Behav. Hum. Perform.* 13:414–32

Steenkamp JBEM, Maydeu-Olivares A. 2021. An updated paradigm for evaluating measurement invariance incorporating common method variance and its assessment. *J. Acad. Mark. Sci.* 49:5–29

Thomas KW, Kilmann RH. 1975. The social desirability variable in organizational research: alternative explanation for reported findings. *Acad. Manag. J.* 18:741–52

Thoresen CJ, Kaplan SA. Barsky AP, Warren CR, de Chermont K. 2003. The affective underpinnings of job perceptions and attitudes: a meta-analytic review and integration. *Psychol. Bull.* 129:914–45

Tourangeau R, Rasinski K, D'Andrade R. 1991. Attitude structure and belief accessibility. *J. Exp. Soc. Psychol.* 27:48–75

Wang TY, Bansal P. 2012. Social responsibility in new ventures: profiting from a long-term orientation. *Strateg. Manag. J.* 33:1135–53

Weijters B, Baumgartner H. 2012. Misresponse to reversed and negated items in surveys: a review. *J. Mark. Res.* 49:737–47

Weijters B, De Beuckler A, Baumgartner H. 2014. Discriminant validity where there should be none: positioning same-scale items in separated blocks of a questionnaire. *Appl. Psychol. Meas.* 38:450–63

Weijters B, Geuens M, Schillewaert N. 2009. The proximity effect: the role of inter-item distance on reverse-item bias. *Int. J. Res. Mark.* 26:2–12

Weijters B, Geuens M, Schillewaert N. 2010. The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Appl. Psychol. Meas.* 34:105–21

Weijters B, Schillewaert N, Geuens M. 2008. Assessing response styles across modes of data collection. *J. Acad. Mark. Sci.* 36:409–22

Williams LJ, Anderson SE. 1994. An alternative approach to method effects by using latent-variable models: applications in organizational-behavior research. *J. Appl. Psychol.* 79:323–31

Williams LJ, Cote JA, Buckley MR. 1989. Lack of method variance in self-reported affect and perceptions at work: reality or artifact. *J. Appl. Psychol.* 74:462–68

Williams LJ, Gavin MB, Williams ML. 1996. Measurement and nonmeasurement processes with negative affectivity and employee attitudes. *J. Appl. Psychol.* 81:88–101

Williams LJ, Hartman N, Cavazotte F. 2010. Method variance and marker variables: a review and comprehensive CFA marker technique. *Organ. Res. Methods* 13:477–514

Williams LJ, McGonagle AK. 2016. Four research designs and a comprehensive analysis strategy for investigating common method variance with self-report measures using latent variables. *J. Bus. Psychol.* 31:339–59

Wilson V, Srite M, Loiacono E. 2021. The effects of item ordering on reproducibility in information systems online survey research. *Commun. Assoc. Inf. Syst.* 49:41

Wothke W, Browne MW. 1990. The direct-product model for the MTMM matrix parameterized as a 2nd-order factor-analysis model. *Psychometrika* 55:255–62

Yao MH, Xu YJ. 2021. Method bias mechanisms and procedural remedies. *Sociol. Methods Res.* In press. **https://doi.org/10.1177/00491241211043141**

Zhang W, Yuan G, Xue R, Han Y, Taylor JE. 2022. Mitigating common method bias in construction engineering and management research. *J. Constr. Eng. Manag.* 148:0402208

Zeng B, Wen H, Zhang J. 2020. How does the valence of wording affect features of a scale? The method effects in the undergraduate learning burnout scale. *Front. Psychol.* 11:585179

Zhu D, Doan T, Kanjanakan P, Kim PB. 2022. The impact of emotional intelligence on hospitality employees' work outcomes: a systematic and meta-analytical review. *J. Hosp. Mark. Manag.* 31:326–47

Zohar D, Luria G. 2004. Climate as a social-cognitive construction of supervisor safety practices: scripts as proxy of behavior patterns. *J. Appl. Psychol.* 89:322–33