Article

Organizational Research Methods
2023, Vol. 26(3) 409–440

DOI: 10.1177/10944281221075361
journals.sagepub.com/home/orm
Sage

Attitude Toward the Color Blue: An Ideal Marker Variable

Brian K. Miller[1] **and Marcia J. Simmering**[2]

Abstract

Researchers often turn to post hoc statistical techniques to identify common method variance (CMV) in same source data and one viable option is to use a marker variable. The choice of marker variable is important, yet these variables are difficult to find, primarily because they must be theoretically unrelated to study variables but measured in the same way (e.g., perceptual; on a Likert scale). This manuscript uses scale development best practices to create a marker variable—attitude toward the color blue—that can be applied in a wide variety of social science research. Scale reliability and validity are addressed, discriminant validity with other measures that detect CMV is tested, and the Confirmatory Factor Analysis Marker Technique is applied with this scale. An experiment designed to analyze the effect of the placement of the scale in surveys is reported. Recommendations to researchers for use of this new scale to detect CMV are provided.

Keywords

common method variance, ideal marker variable, survey research, confirmatory factor analysis marker technique, attitude toward the color blue

Despite some opinions to the contrary (e.g., Spector, 1987; 1994; 2006; Spector & Brannick, 1995) common method variance (CMV) is of great concern to most researchers who collect cross-sectional, single-source data on multiple variables. Method variance has been defined as "systematic variation in an observed variable due to the method used," (Spector et al., 2019, p. 856), and it becomes CMV when the variance is shared between two or more measures that use the same method. Although method variance can introduce error that is unique to a measure and thus deflates substantive correlations--called uncommon method variance or UMV (Spector et al., 2019; Williams & Brown, 1994)--the primary concern with CMV is that it will artificially inflate bivariate relationships that are measured with the same method, on the same subjects, and at the same time (Podsakoff et al., 2003). Despite a range of *a priori* recommendations aimed at avoiding CMV, such as intermixing

[1]Department of Management, Texas State University, San Marcos, TX, USA
[2]Department of Management, Louisiana Tech University, Ruston, LA, USA

Corresponding Author:
Brian K. Miller, Department of Management, Texas State University, 545 McCoy Hall, San Marcos, TX 78666, USA.
Email: brianmillerphd@txstate.edu

items and guaranteeing respondent anonymity (Podsakoff et al., 2003; Podsakoff & Organ, 1986), these potential remedies have been rarely studied empirically, and the small amount of research that addresses them has shown little in the way of efficacy (Castille et al., 2017; Wall et al., 2015). In contrast, *post hoc* efforts at detecting CMV have been studied more extensively.

Recently, the use of an ideal marker variable applied in confirmatory factor analysis has been popularized (Simmering et al., 2015; Williams & Anderson, 1994; Williams & O'Boyle, 2015; Williams et al., 2010). An ideal marker variable is theoretically unrelated to all other variables in a structural equation model, and if analysis shows a relation between the marker and substantive variables, this can be construed to be an artifact of CMV (Richardson et al., 2009). The search for a variable uncorrelated with other variables in social science research has been difficult, but recent reviews of these efforts (Simmering et al., 2015) suggest that items designed to measure attitude toward the color blue (ATCB: Miller & Chiodo, 2008) have some merit. The purpose of the current study is to refine the ATCB scale, examine its psychometric properties in relation to other variables, and demonstrate its use in detecting CMV.

Remedies for CMV

Management and organizational behavior journals have put increasing pressure on researchers to address CMV in same-source data (e.g., Ashkanasy, 2008; Chang et al., 2010). Although post hoc statistical tests of CMV have been studied with some evidence of efficacy (Richardson, et al., 2009), many scholars promote the use of a priori procedural approaches to managing CMV (i.e., temporal separation of measures, distinct source data) over post hoc statistical techniques (Podsakoff et al., 2003; Podsakoaff et al., 2012). Yet, there are three compelling reasons that post hoc statistical remedies for CMV should not be dismissed, and indeed, should be more thoroughly investigated and used.

First, some social science research variables are most effectively studied using a single source, and the use of distinct source data may not improve internal validity. As Chan (2009, p. 310) notes, self-report data, "… although often imperfect, are not inherently flawed." Chan (2009) further argues that a wide array of self-reported constructs in management and psychology have substantial convergent and discriminant validity. Indeed, many self-reported attitudinal and behavioral constructs demonstrate construct validity, as noted by Chan (2009), such as the Big Five personality traits, self-efficacy, and perceived organizational support. Further, there is no compelling evidence that distinct source data is necessarily free from CMV.

Second, there is little empirical evidence that procedural remedies such as temporal separation of data collection, psychological separation of measures on a survey, or the use of distinct source data indeed reduce CMV in data. Despite Podsakoff et al. (2003) pointing to results of multitrait-multimethod matrices (MTMM, Campbell & Fiske, 1959) as providing adequate confirmation of the inflationary nature of CMV, a deeper look at the research they cite indicates a less convincing story. Specifically, in many of the MTMM results, the "other" source has data that are measured very differently from the same variables collected from the single source. For instance, as evidence of CMV Podsakoff et al. (2003) cite the results of a meta-analysis conducted by Fuller et al. (1996), in which same-source and distinct-source corrected correlations among the same constructs differed. Yet, Fuller et al. (1996) noted that the data from the other source were objective data rather than perceptual; that is, the measurement of the constructs was confounded by method. Scant evidence of the efficacy of procedural remedies to reduce CMV exists. Two systematic examinations of a priori CMV reduction through means such as instructional manipulation (Wall et al., 2015), psychological separation, and temporal separation (Castille et al., 2017) found that data collected using such safeguards produced results that were not meaningfully different from data collected without them. Thus, in the absence of further findings, one cannot conclude that preventative CMV measures are effective.

Third, there is evidence that at least one post hoc statistical approach to identifying CMV—the Confirmatory Factor Analysis (CFA) Marker Technique (Williams et al., 2010)—can detect CMV in data at a higher rate of efficacy than other post hoc tests. In an analysis of large-scale simulated data, Richardson et al., (2009) determined that the CFA Marker Technique could be a useful way to identify the presence of CMV in same-source data. Further, Williams and O'Boyle (2015) found that this approach can also identify bias and "… in most practical settings the CFA Marker Technique yields parameter estimates close to their true values …" (p. 1,579). Arguably as important, this technique was highly unlikely to identify CMV when it was not actually present (Richardson et al., 2009). Procedural remedies have not been subjected to the same level of rigorous testing, so empirical evidence indicates that the CFA Marker Variable approach may be the most appropriate way to identify CMV in same source data.

Marker Variables

The marker variable was introduced by Lindell and Whitney (2001) as a way to capture CMV. Their notion was that, if a variable that is measured in the same way as substantive study variables (e.g., on a Likert scale), is perceptual, and is theoretically unrelated to the substantive variables in a study, any relation between the marker and substantive variables could be reasonably determined to be method variance. The Lindell and Whitney (2001) technique used a partial correlation approach, in which the smallest correlation between any substantive variable and the marker variable was partialled out to account for CMV.

Conceptually, when the marker variable is theoretically unrelated to the study variables, it should account for CMV related to response tendencies that are affect-driven, without measuring the affective response directly (Williams & McGonagle, 2016). Podsakoff et al. (2003) identified a variety of sources of CMV, including common rater effects, item characteristic effects, item context effects, and measurement context effects. Marker variables can capture common rater effects, such as mood state (e.g., positive and negative affect), transient mood, the consistency motif (in which respondents attempt to provide consistent answers across items, regardless of content), and implicit theories (when respondents link survey concepts in their minds, and answer based on those connections). Because a marker variable is perceptual and is measured in the same way as study variables (e.g., on a Likert scale), such common rater effects should be captured by it.

Williams et al. (2010) expanded on the notion of marker variable detection by developing the CFA Marker Technique, in which the marker is used in a series of CFA models to determine the degree to which CMV is present and/or contaminates the data. In their empirical analysis of the accuracy of post hoc statistical techniques to detect CMV, Richardson et al. (2009) found that the correlational marker technique and the unmeasured latent method construct approach (in which a latent construct with no unique observed indicators is added to a model with the assumption that it captures method variance; Williams et al., 1989) produced inaccurate findings under most conditions. Further, Chin et al.'s (2012) data simulation found similarly dismal accuracy rates for the unmeasured latent method construct technique.

Although Richardson et al. (2009) found that the CFA Marker Technique was more efficacious than the other post hoc approaches that they tested, there are two primary caveats from their findings. First, they concluded that the CFA Marker Technique worked only when an ideal marker was used. Notably, Richardson et al. (2009) and Williams and O'Boyle (2015) tested only one kind of nonideal marker variable—one that was theoretically related to the substantive variables—and not non-perceptual nonideal marker variables (e.g., age, tenure, firm size). As was found by Simmering et al. (2015), 20.35% of articles that employed the CFA Marker Technique used a nonnideal marker variable that was perceptual or objective (e.g., a demographic variable), as compared to 8.79% that used a nonideal marker variable that was perceptual, but theoretically related to the study variables. Thus, the conclusions found by Richardson et al. (2009) and Williams and O'Boyle (2015) regarding the value of nonideal marker variables only applies to one type of these variables.

A second caveat is that Richardson et al. (2009) concluded that the CFA Marker Technique was accurate in detecting CMV and not necessarily bias. Williams and O'Boyle (2015), however, demonstrated that the CFA Marker Technique could "yield parameter estimates close to their true values" (p. 1,579). Despite this difference in findings, these studies, along with others (e.g., Williams & McGonagle, 2016) continue to demonstrate that the CFA Marker Technique is the most valid post hoc CMV detection technique currently available.

Unlike other popular post hoc correction methods, the CFA Marker Technique has received empirical support (Richardson et al., 2009; Williams et al., 2010; Williams & O'Boyle, 2015), and the use of this technique in published research has increased substantially since its introduction (Simmering, et al., 2015). Richardson et al. (2009) found, however, that the CFA Marker Technique was accurate for detecting CMV, but typically only when an "ideal" marker—a term they adopted to describe Lindell and Whitney's (2001) original marker variable description—was used. Non-ideal marker variables, therefore, were not measured in the same way as the substantive variables, were not perceptual, or were theoretically related to the study variables, and were found to not be useful for identifying CMV (Richardson et al., 2009).

The use of the CFA Marker Technique was limited early on, with Simmering et al. (2015) finding that only 62 articles used it in the three years between 2010 (the year of its introduction) and 2012. A search of articles published only in 2020 that cite Simmering et al. (2015) and/or Williams et al. (2010) conducted by the current authors found that the CFA Marker Technique was used 46 times. Thus, it seems implementation of this technique has grown. Yet, one impediment to its use is that a marker variable must be chosen and included in a survey a priori. We contend that if authors can more easily identify an ideal marker variable that can be used in a diverse set of theoretical models, then they may be more likely to use this efficacious technique, rather than relying on post hoc detection methods that are limited in accuracy.

We further argue that one of the impediments to the use of the CFA Marker Technique is its complexity. Although Williams et al. (2010) outline a process for authors to follow, those who are not highly familiar with structural equation modeling software or the accompanying complex code may find the process daunting. Thus, we have made the Mplus code of how to implement the CFA Marker Technique available to readers.[1]

In Search of an Ideal Marker Variable

In the past decade, scholars have increasingly used marker variables to detect CMV in research (Simmering et al., 2015). Yet, questions still remain as to the choice of marker variable (Simmering et al., 2015; Williams & O'Boyle, 2015). A primary issue is the degree to which the "idealness" of the marker variable matters. Recall that an ideal marker variable is one that is both theoretically unrelated to the study variables, is perceptual, and is measured in the same way as the study variable (e.g., on a Likert scale). Conceptually, the function of the marker variable is to capture response tendencies (e.g., yea-saying, implicit theories) and other common rater effects, and non-perceptual non-ideal marker variables cannot do this. Responses to variables such as shoe size (Podsakoff et al., 2003) or job tenure (Simmering et al., 2015) are highly unlikely to vary in the same way that a survey taker would respond to items measuring attitudes or behaviors. Thus, from a purely theoretical standpoint, non-perceptual marker variables should be rejected. While there is some evidence that a non-perceptual non-ideal marker variable produces results that detect CMV in data (e.g., Simmering et al., 2015 detected CMV with job tenure), this finding may be misleading. Here, two things could be occurring—the non-ideal non-perceptual marker variables could either have a theoretical relationship with substantive variables (i.e., as indicated by the correlation between organizational tenure and overclaiming found in Simmering et al., 2015), or they could be capturing a small amount of error commensurate with, but not caused by, CMV.

Considering classical test theory, which posits that observed relations among variables comprise true and error variance, and that error variance can be produced by a wide variety of influences, then it is possible that a non-ideal non-perceptual marker variable produces a small correlation with substantive variables which is error that might be similar in magnitude to CMV. Evidence of this latter possibility is bolstered by findings by Lance et al. (2010) and Fuller et al. (2016), who found that unreliable measures can attenuate correlations among substantive measures to the degree that it offsets any inflationary CMV. Further complicating the issue are the attenuating effects of uncommon method variance which can also attenuate observed correlations (Williams & Brown, 1994), especially when it arises in different source data (Spector et al., 2019). In other words, small amounts of both inflating and deflating error occur in relationships for a variety of reasons, so identifying such error with a nonideal marker variable could produce accurate results, but not in a theoretically sound fashion.

Although researchers can feel confident rejecting non-perceptual non-ideal marker variables, many may be in a position to use a perceptual non-ideal marker variable. If researchers did not build an ideal marker variable into their data collection, they may attempt to use the CFA Marker Technique with a variable from their data that was not used as a substantive variable in the model. This is inadvisable, and both empirical findings and conceptual arguments discourage the use of a perceptual, non-ideal marker. In their data simulation, Richardson et al. (2009) modeled two non-ideal perceptual marker variables (with 0.20 and 0.40 true correlations with substantive variables). They found that in the CFA Marker Technique was "significantly more accurate with an ideal marker than with a non-ideal marker" (p. 788). Williams and O'Boyle (2015) replicated parts of Richardson et al. (2009) and specified the same level of true correlation of marker variables with substantive variables. Their findings also support the claim that an ideal marker variable produces accurate estimates of substantive relations regardless of the level of CMV in data. Further, their results found that when the marker variable was perceptual, but did not contribute to shared CMV, the findings mirrored those of Richardson et al. (2009) in that the technique produced inaccurate conclusions about the presence of CMV. Yet, in practice, the use of a non-ideal perceptual marker variable may produce accurate estimates of substantive relationships. Richardson et al. (2009) argue that this is due to the CFA Marker Approach accounting for measurement error and "reaching the right conclusion, but for the wrong reason" (p. 786).

This, then, warrants consideration of the conceptual rationale against the use of a theoretically related marker variable. When a researcher uses the CFA Marker Technique with a theoretically related marker variable, it may capture true variance and therefore overestimate CMV. This notion is similar to the concern that using "measured cause" variable for CMV (e.g., negative affect, social desirability) may capture substantive variance in addition to method variance (Simmering et al., 2015; Spector, 2006).

Although inquiry into the function of non-ideal marker variables could be illuminating, a more useful area of study would be to address ideal marker variables, as evidence consistently indicates that ideal marker variables in the CFA Marker Variable Technique produce the most efficacious post hoc recognition of CMV in data. Yet, articles that summarize the use of post hoc testing CMV testing indicate that there is still confusion about the type of marker variable that should be used. One of the major challenges with the choice of marker variable is that it is highly dependent on the variables in the substantive model; a marker variable that is measured similarly to study variables and also theoretically unrelated to them is different in nearly every study. Thus, an ideal marker variable that could be used in a wide variety of social sciences data collections would be highly valuable.

In the search for an ideal marker variable with broad applicability in research, it is important to consider the conceptual rationale behind the use of a marker variable. There are a number of potential sources of CMV and an analysis of responses to a marker variable placed on a same-source survey with substantive variables should account for some degree of these. In particular, the common rater effects of consistency motif, implicit theories (or illusory correlations), social desirability, leniency and acquiescence biases, and mood state should be captured by an ideal marker variable. For instance, a participant answering items with a leniency bias is likely to endorse the higher end of the scale for both substantive and marker variables. Addressing common rater effects in same source data is often more challenging than addressing the other potential sources of CMV identified by Podsakoff et al. (2003): item characteristic effects (e.g., the use of common scale formats), item context effects (e.g., order of items), and measurement context effects (e.g., cross-sectional measure of both predictor and criterion). Thus, the post hoc approach of the CFA Marker Technique using an ideal marker variable is attractive, as there are few other viable ways to capture these sources of CMV.

To further explore the notion of the need for a generic ideal marker variable that could be used in a variety of research studies, we conducted a Google Scholar search for refereed journal articles from 2020 citing Williams et al. (2010) and/or Simmering et al. (2015), as these articles are those most likely to be cited when using the CFA Marker Technique. We identified 46 articles that used the CFA Marker Technique and coded them on the marker variable used. Of the 46 articles, five used a non-perceptual nonideal marker, including respondent job function, respondent age, or business location. Six articles employed a theoretically related, perceptual nonideal marker variable, such as uncertainty or satisfaction with the CEO. The majority of articles in this sample (25) purported to use an ideal marker variable that was both perceptual and not theoretically related to the study variables. However, there could be dispute as to the degree to which these were truly non-related. For instance, Osakwe et al. (2020), in a study of brand values and identity, used access to financial resources as a marker variable, which could influence perceptions of brands. Additionally, Randrianasolo et al. (2020) followed Lindell and Whitney (2001) in the use of satisfaction with life as a marker variable in a study of cultural intelligence and shopping behavior. However, given the affect laden nature of most forms of satisfaction, this marker variable could capture some substantive variance in their model.

Despite the limited nature of this review, it produced a number of interesting marker variables that were not theoretically related to study variables, but that could introduce unintended variance or be off-putting to respondents. These included depression, neighborhood safety perceptions, political self-efficacy, negative cognitions, sensitivity to terrorism, fashion consciousness, variety seeking behaviors, and attitude toward social media use. The introduction of a validated measured intended specifically for use as a marker variable, which does not carry the potential to capture unexplained variance in most work-related substantive variables can be valuable.

Only one variable has been created to specifically serve as an ideal marker variable—Attitude Toward the Color Blue—developed by Miller and Chiodo (2008). This variable was created to be measured on a Likert scale, is perceptual in nature, and is highly unlikely to be theoretically related to most social science variables. Attitudes towards a color are potentially less likely than other attitudes measured as markers (e.g., patriotism) to change based on societal factors, experiences, time etc. Although the original ATCB scale has been used in a few studies, the scale was not developed with the steps indicated by best practices as recommended by Hinkin (1998). Thus, the current study follows such best practices to develop a new, validated ATCB scale.

Study One

Attitude towards a color, as reported by Miller and Chiodo (2008) in an unpublished study, was deemed a good reference point for a marker variable. Rather than assessing attitude towards a person or object, which could be substantively related to other model variables, perceptions about colors were anticipated to be vaguer and therefore more likely to capture response bias from common rater effects that create method variance. For example, a variable measuring general attitudes toward authority figures would be nonideal, as it could be theoretically related to supervisor satisfaction, and a marker measuring attitude toward designer goods could be related to pay satisfaction. In particular, ATCB as a marker variable is likely to capture consistent responding, implicit theories, yea-saying/nay-saying, and social desirable responding, all respondent characteristics that can lead to CMV. Further, ATCB is likely to tap into a level of generality in which items tap into a broader attitude (Tourangeau et al., 2000), which is more likely to address a generic response tendency that causes CMV.

As a conceptual frame for the development of an ideal marker, we sought a generalized attitude that went beyond traditional affective measures, which may or may not capture substantive variance (Spector, 2006). Although the study of attitudes has not produced a consistent result, there is some belief that many attitudes are malleable in part due to context (Wilson & Hodges, 1992). An ideal marker variable that measures an attitude should capture common rater effects elicited from the particular survey items on a single survey as an effective means to identify common method variance particular to that study. We believe that items related to feelings about a certain color are more likely to be malleable, or less "crystalized" than other attitudinal items might be. A traditional approach to understanding survey response (Tourangeau et al., 2000) indicates that respondents go through the stages of comprehension (e.g., paying attention to the instructions and questions), retrieval (e.g., remembering, and perhaps filling in missing details from a question), judgment (e.g., drawing inferences), and response (e.g., mapping judgement onto response category). A crystalized attitude is one that is fairly established and easy to recall from memory that tends to be stable (Schwarz & Bohner, 2001) and is retrieved by "looking up" in memory (the so-called "file-drawer" model of attitudes; Wilson & Hodges, 1992). They are more likely to come from strong opinions or in regards to specific events (e.g., attitude towards current supervisor). Conversely, we argue that attitudes towards a color, rather than towards a person or object, are less likely to be crystalized and could more easily capture CMV because they are more malleable in the moment. Indeed, it seems unlikely that most adults would have a strong, consistent, unchanging opinion about specific colors most of the time. Thus, when asked about attitudes towards the color blue, responses are formed in the moment.

An ideal marker variable has no theoretical relation to substantive variables in a study; to that end, it should also have no overtly positive or negative connotation associated with it. Some colors have affective associations with them (e.g., red can imply anger, yellow may signify cowardice), so Study One sought to verify that the color blue has no overt exclusively and consistently positive or negative connotations associated with it, as compared to other colors. We conducted pilot testing at two public universities in the southern U.S. Twenty-seven MBA students and 16 undergraduate business students responded for class credit. Participants were presented with a Power Point slide show whose first slide read, "Write the first things that comes to your mind when you hear…" and was followed by a numbered list of colors on the second slide presented one at a time for 10 s each: 1. Brown, 2. Red, 3. Orange, 4. Gold, 5. Yellow, 6. Green, 7. Blue, 8. Purple, 9. White, 10. Black, 11. Grey, 12. Maroon. Students were given a sheet of paper to write the first one or two words that came to mind for each color; this activity was anonymous. Some colors had primarily positive words associated with them (e.g., gold was associated with jewelry, the Olympics), but others had more negative words associated with them (e.g., grey was associated with gloom or bleakness). The color blue had a number of neutral words associated with it, such as ocean, sky, or water. Only four participants listed "sad" or "sadness" as a word associated with blue. None listed anything masculine associated with blue. With these qualitative results in mind and despite the passage of time in favor of more modern expressions, it is still possible that some persons are familiar with the expression of "feeling blue" as indicating sadness. Based on these results, the color blue was determined to have the greatest variance of associations attached to it, without any particular skew towards positive or negative.

Study Two

Method and Results

Item Generation Procedures. To determine evidence of content validity, a list of possible items for the scale was created, beginning with Miller and Chiodo's (2008) original eight ATCB items. The current authors individually brainstormed other items for the scale, and solicited individually brainstormed items from other researchers (two PhDs and one doctoral student) familiar with the research project. The resulting list of sixty-nine items was then reviewed by the study authors. First, all redundant items were removed. Second, any items that referenced a physical object (e.g., a blue car) or a place (e.g., blue décor) were discarded due to possible participant reactions to an entity other than the color blue as is problematic with double-barreled items. All items referencing a particular mood, feeling, or psychological state (e.g., "The color blue makes me smile") and any item referring to anything prototypically masculine was eliminated as well (e.g., "Blue is a masculine color"). Lastly, negatively keyed items were removed because of well-known issues associated with the impact of such items on the factor structure of scales (Dalal & Carter, 2015). The list of 69 was trimmed to 14 items. Instructions for these items read, "Please consider your thoughts about the color blue and respond below. There are no right or wrong answers. Some of these items may seem similar to one another, but the repetition is necessary for proper statistical analysis of these items." Responses were on a 7-point Likert scale in which 1 = strongly disagree, 2 = disagree, 3 = somewhat disagree, 4 = neither agree nor disagree, 5 = somewhat agree, 6 = agree, and 7 = strongly agree.

Participants

Responses to these 14 items were then gathered from 417 Amazon.com Mechanical Turk (MTurk) workers in an anonymous online survey. Of those, 42 persons failed to respond to one item and 6 others skipped two items, thus 379 provided complete data. The mean age was 35.06 with a range from 20 to 68 years old. The sample was 56% male and the self-reported race or ethnicity was 63.2% White, 5% Black, 2.9% Hispanic White, 1.6% Hispanic non-White, 22% Asian, 1.1% American Indian, and 4.2% other. The means for other self-reported demographics were: (a) full-time work experience = 11.57 years, (b) part-time work experience = 3.51 years, and (c) current job tenure = 5.10 years. Of the 379 respondents, 311 were currently employed. Of those, 261 were employed full-time and 99 were the manager or supervisor of other workers with mean number of direct reports = 8.99. The distribution of highest achieved educational level was as follows: 0.3% less than high school diploma, 8.2% with high school diploma, 20.8% with some college, 12.9% with a two-year college degree, 41.7% with a four-year degree, 15% with a professional degree, and 1.1% with a doctorate. This sample was randomly divided into two equal halves. Subsample A had 189 respondents whose responses were submitted to exploratory factor analysis and the responses of subsample B ($n = 190$) were used in confirmatory factor analysis.

Subsample A Procedure

To avoid concerns with favored extraction practices (Thompson & Daniel, 1996), we used PCA, principal axis factoring (PAF: a true EFA technique), and EFA with a maximum likelihood extraction. Given the many variables that can affect PCA/EFA results (e.g., non-sphericity, communalities, sample size, rotation method), the decision of an appropriate cut score for factor loadings was of concern. Statistical cut scores of all types have been mythologized (Vandenberg, 2006) and even eulogized (Cumming, 2012; Trafimow & Marks, 2015) but it is widely accepted (e.g., Comrey, 1973; Gorsuch, 1983; Tabachnik & Fidell, 2007; Thompson, 2004) that factor loadings should be at least 0.30 to be minimally acceptable, above 0.40 to be considered important, and above 0.50

Table 1. Preliminary Items and Descriptive Statistics in Study Two (Both Subsamples A and B Combined).

	Mean	SD	Skewness	Kurtosis
Q1. Blue is a beautiful color.	6.06	0.80	−0.97	2.07
Q3. Blue is one of my favorite colors.	5.68	1.30	−1.11	0.98
Q7. I prefer blue to all other colors.	4.19	1.81	−0.14	−1.05
Q8. Blue is a lovely color.	5.92	0.99	−1.12	1.61
Q10. Blue is a pleasant color.	6.04	0.91	−1.18	2.16
Q11. In regards to primary colors (red, blue, and yellow), the color blue is the most beautiful.	5.32	1.60	−0.77	−0.25
Q12. Blue is my favorite color.	4.65	1.88	−0.42	−1.10
Q15. The color blue is wonderful.	5.80	0.96	−0.76	0.71
Q19. I prefer blue to most other colors.	5.01	1.56	−0.63	−0.19
Q23. I love the color blue.	5.60	1.23	−1.05	1.41
Q24. Blue is a nice color.	6.12	0.86	−1.08	1.83
Q26. I think blue is a pretty color.	5.96	0.93	−0.83	0.74
Q27. I like the color blue.	6.03	0.93	−1.41	3.59
Q28. I think blue is the best color.	4.67	1.69	−0.36	−0.75

Note: $n = 379$.

to be practically significant (Hair et al., 1998) for samples of the size in the current study. Therefore, items were retained if they resulted in a factor loading on the first factor in excess of 0.40 and loaded lower than that on any other factor. Each of the three extractions used an orthogonal varimax rotation with a goal of simple structure (Thurstone, 1935) in order to "… maximize our ability to interpret the nature of the latent constructs underlying scores on measured variables" (p. 41) which is the result in the overwhelming majority of varimax rotations (Thompson, 2004). Lastly, despite the magnitude of the loading on a factor, strongly cross-loaded items were not retained.

Subsample A Results. The item-level skewness and kurtosis for the 14 items ranged from −1.51 to −0.03 and from −1.136 to 4.70, respectively. These values fall short of the limits established by West et al. (1995) of |2.0| for skewness and |7.0| for kurtosis and the data are therefore univariate normal. See Table 1 for these results. Each item was positively and significantly correlated with every other item at $p < .001$ with correlations ranging in magnitude from 0.29 to 0.79. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.94 and the approximate chi-square for Bartlett's test of sphericity was 2350.69 ($df = 91$, $p < .001$). Hair et al. (1998) suggest that the KMO measure should exceed 0.80 for a "meritorious" (p. 99) prediction of each variable by every other variable. The test of sphericity analyzes the likely significance that a sufficient number of items are significantly correlated so that PCA/EFA may proceed. With these statistics in mind, the data were deemed appropriate for further analysis.

The variance explained by the first component or factor was 60.03% suggesting that the majority of the variance in the items emanated from a common underlying source. For each extraction a two-component or two-factor solution was best according to the criteria of eigen values > 1 and the scree plots. Using the minimum thresholds of factor loadings ≥ 0.40, each extraction method resulted in the same survey items as candidates for retention. The eigen value for the first factor was 8.40 and for the second factor it was 1.98. The first factor appeared to measure favorable attitudes toward the color blue (ATCB) and was comprised of seven items. Cronbach's coefficient alpha of internal consistency reliability for scores on these seven items was 0.94. The second factor seemed to measure a preference for the color blue (PCB: e.g., "I prefer blue to all other colors") and was comprised of four

Table 2. Factor Loadings[a] of Preliminary Items from Study Two Subsample A[b].

	PCA rotated component matrix		PFA rotated factor matrix		ML rotated factor matrix	
Components or factors:	1	2	1	2	1	2
Q1. Blue is a beautiful color.	*0.766*	0.270	*0.727*	0.281	*0.725*	0.282
Q3. Blue is one of my favorite colors.	0.566	0.648	0.568	0.623	0.568	0.613
Q7. I prefer blue to all other colors.	0.119	*0.890*	0.144	*0.850*	0.144	*0.856*
Q8. Blue is a lovely color.	*0.862*	0.168	*0.835*	0.180	*0.832*	0.182
Q10. Blue is a pleasant color.	*0.873*	0.129	*0.847*	0.141	*0.846*	0.140
Q11. In regards to primary colors (red, blue, and yellow), the color blue is the most beautiful.	0.438	0.574	0.441	0.520	0.447	0.510
Q12. Blue is my favorite color.	0.173	*0.901*	0.187	*0.883*	0.18	*0.890*
Q15. The color blue is wonderful.	*0.746*	0.325	*0.713*	0.329	*0.716*	0.329
Q19. I prefer blue to most other colors.	0.350	*0.766*	0.370	*0.714*	0.381	*0.697*
Q23. I love the color blue.	0.632	0.598	0.630	0.581	0.632	0.578
Q24. Blue is a nice color.	*0.811*	0.283	*0.786*	0.289	*0.788*	0.289
Q26. I think blue is a pretty color.	*0.852*	0.257	*0.837*	0.260	*0.836*	0.258
Q27. I like the color blue.	*0.813*	0.315	*0.796*	0.315	*0.803*	0.306
Q28. I think blue is the best color.	0.235	*0.892*	0.246	*0.881*	0.249	*0.886*

[a]Loadings meeting *a priori* thresholds of $\geq$0.40 in italics.
[b]Varimax rotation on each extraction.

cleanly loading items with a Cronbach's alpha of 0.92. Three other items showed large cross-loadings on both factors and were not subjected to further analysis. See Table 2 for the nearly identical results via each extraction method for all 14 items.

Subsample B Procedure. The seven items (items 1, 8, 10, 15, 24, 26, 27) loading on the first factor and the four items (items 7, 12, 19, 28) loading on the second factor in the EFA on Subsample A were submitted to confirmatory factor analysis (CFA) using responses from the independent Subsample B and analyzed with LISREL 8.80 software (Jöreskog & Sörbom, 2006). The skewness and kurtosis of these 11 items ranged from −1.27 to −1.00 and from 1.57 to 3.25, respectively. However, univariate normality is only a necessary but not sufficient condition for multivariate normality (Henson, 1999) so Mardia's coefficient of multivariate kurtosis was calculated using the DeCarlo (1997) macro. The normalized (i.e., standardized) coefficient was 163.55. Using the cutoff of |3.0| recommended by Bentler (1998) and Bentler and Wu (2002) the data were determined to not be multivariate normal. Thus, the Satorra-Bentler scaled chi-square and robust standard errors correction to the maximum likelihood estimation procedure (Satorra & Bentler, 2001) was used in the analysis. Input data consisted of the covariance and asymptotic covariance matrices. Error variances were not allowed to correlate. The metric was set by constraining the single latent construct variances to 1 so that all items loadings were freely estimated and forced onto their respective constructs (ATCB and PCB). Additionally, the covariance between the two factors was freely estimated.

Subsample B Results. These results confirm those of the EFA on Subsample A. The two-factor solution for the 11 items was acceptable in this subsample. The Satorra-Bentler scaled chi-square (S-B χ^2) for the model was 77.96 ($df = 43$, $p < .001$). The supplementary fit indices were as follows: CFI = 0.988, SRMR = 0.065, and RMSEA = 0.066 (90% C.I.: 0.042, 0.089). The correlation between

Table 3. Fit Indices for One-factor Models of Attitude Toward the Color Blue Scale Items.

	Sample size	McDonald's ω	S-B χ^2	CFI	SRMR	RMSEA (90% C.I.)
Study two	379	0.93	24.585*	0.989	0.023	0.045 (0.010, 0.073)
Study three	398	0.96	30.090**	0.988	0.016	0.054 (0.027, 0.080)
Study four	771	0.93	13.373 (*ns*)	1.000	0.013	0.000 (0.000, 0.033)
Study five	609	0.91	15.862 (*ns*)	0.998	0.008	0.014 (0.000, 0.042)

Note: Degrees of freedom for each model are 14. S-B χ^2 = Satorra-Bentler scaled chi square and robust standard error correction, *ns* = non-significant, CFI = comparative fit index, SRMR = standardized root mean squared residual, RMSEA = root mean square error of approximation, CI = confidence interval, ω = omega, a measure of internal consistency reliability.
*$p < .05$.
**$p < .01$.

the ATCB and PCB factors was 0.476 suggesting that the two constructs were positively related but not so strongly correlated as to suggest collinearity (Kline, 2011).

With these results in mind, a within-subjects comparison of the ATCB and the PCB scale using Subsample B followed. The ATCB scale had the following: mean = 5.96, standard deviation = 0.76, skewness = −0.97, and kurtosis = 1.91. These statistics for the PCB scale were 4.70, 1.61, −0.40, and −0.75, respectively. The range of responses for ATCB was 2.86 to 7.00 and for PCB it was 1.00 to 7.00. A comparison of the means resulted in $t = 12.29$ ($df = 188$, $p < .001$). Overall, responses to the ATCB scale were stronger, more narrowly distributed, more left-skewed, and more leptokurtotic than the responses to the PCB scale, within subjects. Additionally, in post hoc analysis a one-factor CFA model of the seven ATCB items in the combined subsamples ($n = 379$) resulted in excellent fit with S-B $\chi^2 = 24.585$, $p < .05$, CFI = 0.989, SRMR = 0.023, and RMSEA = 0.045 (90% C.I.: 0.010, 0.073). See Table 3 for the one-factor fit indices.

Discussion

Study Two followed best practices for scale development to create two distinct scales: *attitude toward the color blue* and *preference for the color blue*. Each scale demonstrated internal consistency reliability, content validity, and discriminant validity in regard to each other. Yet, one question regarding the usefulness of marker variables is the degree to which they are distinguishable from other variables that could be used to detect CMV. Thus, in Study Three, evidence of discriminant validity is further addressed.

Study Three

As a means to begin to establish the construct validity of ATCB and PCB, their distinctiveness must be assessed. Discriminant validity is indicated when the variable of interest does not correlate with those for which it should be dissimilar (Hinkin, 1998). Prior research has recommended that the use of variables that are likely to capture common rater effects, like positive affectivity and social desirability, can address CMV (Podsakoff et al., 2012; Simmering et al., 2015). Yet, Spector et al. (2019) caution that measures such as social desirability may not capture CMV in some variables or may be related in a meaningfully conceptual way to others.

For a marker variable to be of use above and beyond such scales, it must not be so highly correlated with such variables that their use is redundant. Thus, in Study Three, we examined the nature of the relationship between ATCB and PCB with other variables often associated with CMV in order to collect evidence of discriminant validity.

Hypothesis 1: ATCB and PCB will demonstrate discriminant validity from one another.

Despite traditional scale development emphasizing both convergent and discriminant validity (Hinkin, 1998), the current article is intended to create a variable that measures response tendencies differently from other scales. Thus, our primary interest is to demonstrate discriminant validity from (a) other potential maker variables and (b) variables that capture constructs that might be specifically related to the color blue. The Neutral Objects Satisfaction Questionnaire (NOSQ; Weitz, 1952) has been used previously as a marker variable (Johnson et al., 2011), and could be construed as an effective generic marker variable, as it is likely to capture a general attitude that could influence common rater effects.[2] However, in the current investigation, we posit that ATCB and PCB will be distinct from the NOSQ, due to many of the items in the NOSQ having a meaning today that was not likely intended when it was written in the 1950s. Specifically, the items asking about *local newspapers* and *public transportation* are likely to be irrelevant to young or rural respondents. Additionally, several of the scale items could be more politically charged in the United States today than in prior years: asking about the *climate where you live* could be related to perceptions regarding climate change, which differ among Americans; *advertising* could be linked to concerns about artificial intelligence determining ads on social media and beliefs about large technology firms; perceptions of the *local newspaper* could be linked to beliefs about the media overall, and some political groups' current mistrust of many media outlets; *today's cars* may relate to attitudes about fossil fuels, manufacturing, and sustainability; and even *your first name* could be related to racial and ethnic identity. Undoubtedly, political and cultural attitudes could be linked to variety of work-related perceptions. Thus, although this scale surely could be used as an ideal marker variable successfully in some studies, we do not believe that many NOSQ items are as neutral as they were in the 1950s. Unlike the NOSQ, the attitudes and preferences for the color blue are less likely to garner opinions in this way, and therefore, we hypothesize:

Hypothesis 2: ATCB and PCB will exhibit discriminant validity with NOSQ.

In Study 1, we used a qualitative approach to determine any concepts that could be related to the color blue that might also indicate a particular influence in workplace attitudes and behaviors. Although none of our respondents linked the color blue with the concept of masculinity, this is still a conceptual possibility. Thus, to verify that masculine gender role identity was not conflated with items about the color blue, we posited:

Hypothesis 3: Masculine gender role identity will demonstrate discriminant validity with ATCB and PCB.

Finally, to be useful as an ideal marker variable, ATCB/PCB must not simply measure response tendencies that can be captured by existing measures. Termed "presumed CMV cause variables" (Simmering et al., 2015), positive affectivity, negative affectivity and social desirability may lead to CMV (Podsakoff et al., 2003). Due to the conceptual nature of ATCB and PCB, it is possible that scores on these scales reflect a more generalized state of mind, such as affectivity. Similarly, respondents who engage in socially desirable responding (SDR) to substantive items are likely to do the same with marker variables. The impact of SDR is of concern because it can impact self-reports on other scales and measures of SDR can also be used as a marker variable (Simmering et al., 2015) because they capture response tendencies. Thus, distinguishing ATCB and PCB from these variables is critical to these new scales' usefulness.

Hypothesis 4: ATCB and PCB will demonstrate discriminant validity with (a) positive and negative affectivity and (b) social desirability.

Method

Participants. In Study Three, data were collected with an online self-report survey administered to 500 Amazon.com MTurk workers. All data were cross-sectional and anonymous. Of the 500 respondents, the mean age was 34.45 with a range from 20 to 71 years old. The sample was 58.9% male and the self-reported race or ethnicity was 73.6% White, 8.2% Black, 3.4% Hispanic White, 1.4% Hispanic non-White, 11.0% Asian, 1.8% American Indian, and 0.6% other. The means for other self-reported demographics were: (a) full-time work experience = 12.81 years, (b) part-time work experience = 2.84 years, and (c) current job tenure = 5.68 years (all were currently employed). Of those, 434 were employed full-time and 148 were the manager or supervisor of other workers with mean number of direct reports = 10.21. The distribution of highest achieved educational level was as follows: 0.6% less than high school diploma, 10.0% with high school diploma, 21.2% with some college, 12.0% with a two-year college degree, 43.3% with a four-year degree, 11.6% with a professional degree, and 1.2% with a doctorate.

It is of note that 102 respondents failed to respond to one or more items. Therefore, we conducted an attrition analysis comparing the demographic characteristics and available scale scores of those who completed the survey ($n = 398$) to those who skipped one or more items ($n = 102$). Chi-square tests for categorical variables (e.g., gender) and t-tests for continuously scored variables (e.g., age) found no significant differences for respondents completing the demographic variables between the two groups.

Among the scale scores described below, only Negative Affectivity (NA) showed significantly different mean scores for the two groups with $t = 2.69$ ($df = 126.04$, $p < .01$). Those who omitted one or more items other than NA but who did complete the Negative Affectivity items had higher scores ($\text{mean}_1 = 2.01$) than those who completed the entire study ($\text{mean}_2 = 1.67$). Only Negative Affectivity and Impression Management differed on their variance according to Levene's test of the homogeneity of variance. For IM the variance (s^2) for the group who failed to complete other non-IM items in the survey ($n = 87$) was 0.72 and the $s^2 = 1.03$ for the group who completed all survey items ($n = 397$). For Negative Affectivity, $s^2 = 1.37$ for those who failed to complete one or more other non-NA items in the survey ($n = 92$) and $s^2 = 0.79$ for those who completed the survey. Therefore, of all measured variables the dropouts differed from those who completed the survey only in their more narrowly distributed IM responses, more widely distributed NA responses, and higher levels of NA.

Measures

Perceptions of Neutral Objects. Following Johnson et al. (2011), we used the 11 items from the original larger Weitz (1952) Neutral Object Satisfaction Questionnaire, as used by Judge and Bretz (1993). This scale uses a three-point response scale where 1 = dissatisfied, 2 = neutral, and 3 = satisfied. Sample items include: "The city in which you live" and "Your relaxation time." Cronbach's alpha was 0.764 for scores on this instrument.

Masculine Gender Role Identity. This construct was measured with the 20-item scale from Bem (1974). Responses were gathered with a seven-point Likert-type scale anchored by 1 = *never or almost never* and 7 = *almost always true*. Respondents rated themselves on items like "Analytical" and "Dominant" regarding how they see themselves and not as others see them. Cronbach's alpha for scores was 0.927

Positive Affectivity. Using a five-point Likert-type scale, responses were gathered on the positive 10-item portion of the Positive Affectivity Negative Affectivity Scale (PANAS: Watson et al., 1988). Respondents were instructed to respond to the Positive Affectivity (PA) items regarding "to the extent to which (they) generally fe(lt) this way, that is, how (they) fe(lt) on average." The response scale was anchored by 1 = *very slightly or not at all* and 5 = *extremely*. Sample items include: "Excited" and "Alert". Cronbach's alpha was 0.927.

Negative Affectivity. This construct was measured with the 10 negative items from the PANAS (Watson et al., 1988). Using the same instructions and response scale as for PA, the NA portion included items like: "Irritable" and "Afraid." Alpha for these scores was 0.959.

Impression Management. This form of socially desirable responding was measured with the 20-item Impression Management (IM) sub-scale of the Balanced Inventory of Desirable Responding (Paulhus, 1988). Data were gathered from respondents using a seven-point Likert-type response scale anchored by 1 = *not true* and 7 = v*ery true*. Sample items include: "I sometimes tell lies if I have to" (reverse scored) and "I always obey laws, even if I'm unlikely to get caught." Alpha for scores on this instrument was 0.851.

Preference for the Color Blue. Using the four items extracted via EFA and confirmed via CFA from Study Two, responses to the PCB scale were gathered in this independent sample. These items used a true Likert scale anchored by 1 = *strongly disagree* and 7 = *strongly agree*. Alpha for scores on the PCB scale was 0.952. Instructions for this scale were, "Please indicate your level of agreement or disagreement with the following statements by using the 1 - 7 response scale below."

Attitude Toward the Color Blue. Using the seven items previously extracted with EFA and confirmed with CFA in Study Two and the same response scale and instructions as for PCB, responses were gathered in this sample. See the Appendix for the items. Alpha was 0.965.

Results

Item Analysis Results. The distributional properties of the items in the ATCB and PCB scales were examined. For the ATCB scale, item skewness ranged from −1.54 to −1.09 and kurtosis ranged from 1.28 to 3.10. The means on the items ranged from 5.58 ("The color blue is wonderful") to 5.83 ("Blue is a beautiful color"). The item's standard deviations ranged from 1.21 to 1.35. Correlations between the seven items ranged from 0.67 ($p < .001$) to 0.85 ($p < .001$).

For the PCB scale, item skewness ranged from −0.51 to −0.13 and kurtosis ranged from −1.36 to −0.82. The means on the items ranged from 4.13 ("I prefer blue to all other colors") to 4.65 ("I prefer blue to most other colors"). The item's standard deviations ranged from 1.85 to 2.15. For the items in both scales the skewness and kurtosis fell short of the cutoffs for non-normality recommended by West et al. (1995) and therefore indicate univariate normality. Correlations between the four PCB items ranged from 0.77 ($p < .001$) to 0.90 ($p < .001$).

In this study, as in the previous study, a within-subjects comparison of the ATCB and the PCB scale was conducted on 476 respondents who provided complete data on both scales. At the scale level, ATCB had a mean of 5.78 and standard deviation of 1.14. The PCB scale had a mean of 4.31 and standard deviation of 1.90. The comparison of the means resulted in $t = 18.66$ ($df = 475$, $p < .001$) such that scores on the ATCB scale were higher than the scores on the PCB scale.

Correlational Analysis Results. To examine the relationships between these measured scales, Pearson correlations and a point biserial correlation for gender's relationship with the other variables were calculated using listwise deletion. Preference for the Color Blue was significantly positively related to four of the six variables in this study. Only IM and gender were not significantly related to PCB. Attitude Toward the Color Blue was only significantly related to two of the six variables: Neutral Objects Scale ($r = 0.151$, $p < .01$) and Positive Affectivity at $r = 0.140$ ($p < .01$). As in the results of Study Two's Subsample B, PCB and ATCB were similarly correlated here at $r = 0.439$ ($p < .001$) in Study Three. All of these relationships were far below the cutoff for collinearity of

Table 4. Study Three Scale Correlations, Means, Standard Deviations, and Reliabilities.

	1	2	3	4	5	6	7	8
1. Gender[a]	--							
2. Neutral objects	−0.063	(0.76)						
3. MGRI[b]	0.281***	0.305***	(0.93)					
4. Positive affectivity	0.050	0.377***	0.615***	(0.93)				
5. Negative affectivity	0.023	−0.100*	−0.007	0.013	(0.96)			
6. Impression management	−0.066	0.133**	0.049	0.210***	−0.182**	(0.85)		
7. Preference for the color blue	0.095	0.153***	0.131**	0.227***	0.169***	0.091	(0.95)	
8. Attitude toward the color blue	−0.075	0.151***	0.069	0.140***	−0.082	−0.016	0.439***	(0.97)
Mean	0.57	2.29	4.72	3.42	1.73	4.07	4.30	5.78
Standard deviation	0.50	0.40	1.03	0.85	0.96	0.99	1.89	1.14

Note: Reliabilities appear on the diagonals in parentheses. $n = 398$.
[a]Coded as 0 = female, 1 = male.
[b]MGRI = masculine gender role identity.
*$p < .05$.
**$p < .01$.
***$p < .001$.

0.85 (Kline, 2011), thus lending support for the discriminant validity of both the ATCB and PCB scales, and thus supporting Hypotheses 1 through 4. See Table 4 for these results. In post hoc analysis the fit of a one-factor CFA model to the seven ATCB items was excellent. See Table 3 for these results.

Discussion

Study Three addressed the discriminant validity of ATCB and PCB and found that both were distinct constructs from other similar variables that have been used to assess CMV. More specifically, ATCB was more weakly related to all six variables than was PCB, indicating that it has stronger discriminant validity. Because of this, analysis of PCB was henceforth discontinued as its distinctiveness from other variables used as marker variables was not as prominent, which reduces its usefulness as a unique marker variable.

Study Four

In Study Four, detection of CMV in a substantive variable relationship using the new ATCB scale was examined. To do this, the commonly studied bivariate relationship between organizational commitment and job satisfaction was estimated. The relationship between these two variables should be highly prone to CMV—they are transparent measures and fairly short scales. Thus, survey takers should be able to easily determine the relationship between variables that the researchers are trying to capture (i.e., a positive one) and also recall their answers to prior items when going through the survey. This should enhance common rater effects that are linked to demand characteristics and memory, such as implicit theories. Further, two short measures on the same Likert scales should also enhance item characteristic effects that can lead to CMV. This data collection therefore produces a rich context in which to test for CMV using ATCB as an ideal marker variable in the CFA Marker Technique.

The other strength of the use of the bivariate relationship between organizational commitment and job satisfaction is that published meta-analytically derived correlations for it exist. Thus, there are benchmarks against which we can measure the correlation that the data collection in Study Five produces. Mathieu and Zajac (1990) reported an uncorrected mean r of 0.488 and Tett and Meyer (1993) reported a mean r of 0.586 that was sample-size weighted but also uncorrected for measurement error. In the event that this relationship in Study Five substantially exceeds the magnitude of its meta-analytic counterparts, ATCB should be able to detect CMV in the relationship.

Hypothesis 5: ATCB will detect CMV in an unusually strong correlation in same-source data.

Method

Self-reports of organizational commitment and job satisfaction were collected from Amazon.com's MTurk workers. Participants were recruited with a description that the survey was seeking attitudes and opinions of people currently working in the U.S. Pay for the survey was $1 for a few minutes of completion time, and 865 people attempted the survey. Of that group, 56 were not allowed to complete the survey because they did not work in the U.S. and 15 did not finish the survey. Data from 24 participants were eliminated through listwise deletion, resulting in 771 usable surveys. One criticism of Mechanical Turk worker data is that respondents such as this might be more susceptible to demand characteristics (Lovett et al., 2018). However, this is more likely to lead to possible inflation via common method variance thus creating an ideal situation in which to test a new marker variable.

The mean age of the sample was 33.64, ranging from 18 to 68 years, and the sample had slightly more males (58.55%) than females. The sample self-reported ethnicity as 67.27% White, 9.48% Black, 3.90% Hispanic White, 1.43% Hispanic Non-White, 11.69% Asian, 2.99% American Indian, and 2.97% other. The average number of years that the sample had with full-time work experience was 10.71, and the average number of years of part-time work experience was 3.60. All participants were currently working, and the average current job tenure was 5.05 years. The number of participants working full-time was 682, and of those 429 were currently supervising others, with an average number of direct reports of 16.84. The highest achieved educational level was as follows: 0.001% less than high school diploma, 8.45% with high school diploma, 16.12% with some college, 8.45% with a two-year college degree, 47.07% with a four-year degree, 17.69% with a professional degree, and 2.08% with a doctorate.

Measures

Organizational Commitment. This construct was measured using the nine-item Shortened Organizational Commitment Questionnaire from Mowday et al. (1979). Sample items include: "I talk up this organization to my friends as a great organization to work for" and "I really care about the fate of this organization". Cronbach's alpha for scores on this instrument was 0.943.

Job Satisfaction. This construct was measured using the five-item scale from Hackman and Oldham (1975). Sample items include: "I am generally satisfied with the kind of work I do in this job" and "People on this job often think of quitting" (reverse scored). Alpha reliability for scores on this scale was 0.795.

Attitude Toward the Color Blue. In this study, ATCB was measured using the previously created and analyzed seven items. Alpha for these scores was 0.936.

Analysis. These data demonstrated univariate normality, based on cutoffs suggested by West et al. (1995). For the organizational commitment scale, item skewness ranged from −1.07 to −0.47, and kurtosis ranged from −0.89 to 0.82. The job satisfaction items had a range of skewness of −1.12 to −0.03 and a range of kurtosis of −1.42 to 0.80. Finally, item skewness for ATCB ranged from −1.53 to −0.93, and item kurtosis ranged from 0.68 to 3.32. However, univariate normality is a necessary but not sufficient condition for multivariate normality (Henson, 1999). By using Mplus 8.2 software (Muthén & Muthén, 2018) it was determined that the data were not multivariate normal as the scaling correction factor was 1.578. For multivariate normality to be insured the correction factor should be 1.0.

Therefore, the Satorra-Bentler corrections to the maximum likelihood method of estimation were implemented. Following the steps outlined by Williams et al. (2010) seven models were tested and compared: CFA, Baseline, Method-C, Method-U, Method-R, Method-S(0.05), and Method-S(0.01). These models were then compared using a chi-square difference test to determine whether CMV was present, and if it was, whether it biased the focal substantive variable relationship between organizational commitment and job satisfaction. The model fit comparisons were not simple difference tests as they required adjustments because of the scaling factor adjustment to the chi-square (Satorra, 2000).

In this first phase of analysis, the models used to conduct the CFA Marker Technique are described below. In the CFA model, the three latent variables were allowed to correlate freely and the items were forced to load onto their respective constructs. From this model, the factor loadings and measurement error variance estimates were then used for subsequent model estimation. In the Baseline model, the correlation between organizational commitment and job satisfaction was allowed to be freely estimated, the correlation between these two variables and the marker variable (ATCB) was constrained to orthogonality (i.e., zero), and the marker variable's indicators were set with the fixed factor loadings and fixed error variances obtained from the CFA Model. A third model, the Method-C Model, was then estimated. This model allowed for the correlation between organizational commitment and job satisfaction to be freely estimated and ATCB to be uncorrelated with either substantive construct, just as in the Baseline Model. However, the Method-C model added factor cross-loadings between ATCB and the two substantive variables' items. Additionally, the factor loadings from ATCB to the items measuring the substantive variables were constrained to be equal. Comparison of the Baseline and Method-C Models indicates the degree to which the substantive relationship is contaminated by method variance captured by the marker variable. To determine the degree to which CVM was biasing, two more models were estimated. The Method-U model is similar to the Method-C Model, except that the factor loadings between the latent marker variable (ATCB) and each of the substantive variables' items were unconstrained and allowed to be freely estimated. A comparison of the Method-U and Method-C models indicates whether there is a difference in the unconstrained and constrained models. The Method-R Model was calculated and compared to the Method-U Model to determine the degree to which CMV actually had biasing effects on the substantive relationship between job satisfaction and organizational commitment. Additionally, the Method-R Model fixed the correlation between the substantive variables to the same latent factor correlations obtained in the Baseline Model.

The second phase of analysis was to decompose the reliability of the latent variables in the preferred model of the five above into substantive reliability and method reliability. This requires the use of the completely standardized factor loadings and error variances from either Method-C or Method-U. Formulae to examine these statistics in Williams et al. (2010) are based upon the original work of Werts et al. (1974).

The third phase of analysis is sensitivity tests based upon the Method-S(0.05) and Method-S(0.01) models that required that paths from the marker variable to the substantive items be fixed to the upper end (alpha = 0.05) and (alpha = 0.01) of the confidence intervals around the factor loadings obtained in either the Method-C or Method-U model, whichever is best supported. The goal of these sensitivity tests was to examine the impact of sampling error on the substantive variable relationship. The constraint of the paths to higher values than obtained in the best fitting model is a test of their effect on the substantive variable correlation.

Results

The observed score correlation between organizational commitment (OC) and job satisfaction (JS) was 0.684 in these data. This is much larger than the meta-analytic sample size weighted average observed score correlations of 0.488 (Mathieu & Zajac, 1990) and 0.589 (Tett & Meyer, 1993). In the current study, the latent correlation between these two focal constructs was 0.920, which exceeds Kline's (2011) minimum standard of |0.85| for collinearity. Additionally, the latent correlations between ATCB with organizational commitment and by job satisfaction were 0.339 ($p < .001$) and 0.346 ($p < .001$), respectively. With these preliminary results in mind and with CMV being likely to be present, the results of the CFA marker technique as outlined by Williams et al. (2010) and detailed above are described below.

In the first phase of the analysis three latent variables were allowed to correlate freely in a CFA model. From this model, the factor loadings and measurement error variance estimates were then used for subsequent model estimation. The fit index results were S-B $\chi^2 = 788.685$ ($df = 186$, $p < .001$), CFI = 0.926, SRMR = 0.053, RMSEA = 0.065 (90% C.I.: 0.060, 0.070). As noted above, the latent correlation between JS and OC was 0.920. This model fit the data well. The Baseline model fit similarly well except for the SRMR was out of range of acceptability. For the comparison of the Baseline Model and the Method-C model the S-B $\Delta\chi^2$ b was 71.484 ($\Delta df = 1$, $p < .001$), indicating that CMV was present. Additionally, the SRMR returned to within the range of acceptability and the latent correlation between the two focal constructs fell slightly to 0.910 in the Method-C model. A comparison of the Method-U and Method-C models resulted in S-B $\Delta\chi^2$ of 37.138 ($\Delta df = 13$, $p < .001$) with the correlation between JS and OC at 0.908 in the Method-U model. The better fitting model was Method-U, meaning that CMV affects the substantive variables' indicators differently (i.e., congeneric CMV). In comparing the Method-R Model to the Method-U Model the S-B $\Delta\chi^2$ was 0.524 ($\Delta df = 1$, *ns*) which indicates that there is no bias emanating from CMV in the substantive variable correlation and the preferred model is Method-U.

In the second phase of analysis, the reliability of the JS and OC constructs was decomposed into method and substantive reliability. From the Method-U model, the total reliability was 0.793 and 0.945. For ATCB it was 0.936. For JS, the decomposed substantive and method reliabilities were 0.704 and 0.088, respectively. For OC, those values were 0.836 and 0.109, respectively. Overall, the percentage of reliability attributable to the marker variable was 11.10% for JS and 11.53% for OC.

The third phase of analysis required sensitivity test model comparisons to the Method-U model. The S-B $\Delta\chi^2$ was 7.484 ($\Delta df = 14$, *ns*) for the Method-S(0.05) model. The S-B $\Delta\chi^2$ was 0.001 ($\Delta df = 14$, *ns*) for the Method-S(0.01) model. These differences indicate that fixing marker variable loadings to the substantive indicators to values at the high end of the 95% and 99% confidence intervals around those paths in the Method-U model had no impact on the interpretation of the correlation between JS and OC which rose only slightly to 0.909 in all three models.

Results of the CFA Marker Technique demonstrate that ATCB can be used successfully as a marker variable to identify the presence of, and bias introduced by, CMV in a correlation in commonly measured constructs. Thus, hypothesis 5 was supported. See Table 5 for all fit indices and difference test statistics in Study Four. In post hoc analysis the fit of a one factor CFA model for the seven ATCB items as a stand-alone measure was excellent again. See Table 3 for these results.

Table 5. Study Four Fit Indices and Model Comparison Tests Using the CFA Marker Technique.

Study Four ($n = 771$)	S-B χ^2	df	SCF	$\Delta\chi^2$	Δdf	CFI	SRMR	RMSEA (90% C.I.)	r_{js-oc}
CFA model	788.685***	186	1.5830	--	--	0.926	0.053	0.065 (0.060, 0.070)	0.920
Baseline model	827.438***	202	1.6165	--	--	0.923	0.144	0.061 (0.056, 0.065)	0.920
Method-C model	774.455***	201	1.6187	71.484***[a]	1[a]	0.929	0.060	0.061 (0.056, 0.065)	0.910
Method-U model	734.343***	188	1.6431	37.138***[b]	13[b]	0.933	0.050	0.061 (0.057, 0.066)	0.908
Method-R model	734.403***	189	1.6443	0.524[c]	1[c]	0.933	0.050	0.061 (0.057, 0.066)	0.920[e]
Method-S(0.05) model	752.257***	202	1.6165	7.484[d]	14[d]	0.932	0.070	0.059 (0.055, 0.064)	0.909
Method-S(0.01) model	736.410***	202	1.6165	0.001[d]	14[d]	0.932	0.082	0.060 (0.055, 0.064)	0.909

Note: $\Delta\chi^2$ values are not simple difference tests and required special calculations (Satorra, 2000). SCF = scaling correction factor included for purposes of reproducibility. r_{js-oc} = correlation between job satisfaction and organizational commitment.

[a] Baseline vs. Method-C.
[b] Method -C vs. Method-U.
[c] Method-U vs. Method-R.
[d] Method-S vs. Method-U.
[e] Fixed parameter.
***$p < .001$.

Discussion

Studies One through Three followed best practices to develop a new scale. In Study Four, the effectiveness of the ATCB scale for use in the CFA Marker Technique was examined. Using a frequently examined relationship—between organizational commitment and job satisfaction—the marker variable of ATCB identified the presence of CMV but found no bias. A strength of the CFA Marker Technique is that it is able to separately test for the presence of CMV and the degree to which it might be biasing (Richardson et al., 2009). Other more commonly used techniques (e.g., Harman's One-Factor Test) do not allow for this (Fuller et al., 2016). Yet, to properly employ the CFA Marker Technique, an ideal marker variable is needed. As shown in Study Four, in a bivariate correlation in which CMV is highly likely, ATCB is capable of identifying the presence of CMV and the degree of its biasing effect. We conclude that ATCB can be used effectively as an ideal marker variable in data such as this.

Study Five

Study Four established the ability of the ATCB scale as a marker variable to detect CMV in an unusually strong bivariate relationship. The purpose of Study Five was to examine the impact of the position of the ATCB items in a survey when using the CFA Marker Technique. Prior research has posited that a marker variable placed between the independent and dependent variables in a study can create a psychological separation that may disrupt response tendencies and reduce CMV (Podsakoff et al., 2003). Yet, other authors argue that there is little evidence as to how placement of attitudinal variables on a survey relates to their susceptibility to context in the rest of the survey (Tourangeau et al., 2000). There has been no empirical comparison of different ideal marker variable locations within a survey. Therefore, Study Five employed three experimental conditions to determine the degree to which marker variable placement affected its use in the CFA Marker Technique. The three conditions were that the ATCB scale was: (1) placed with all ATCB items together between the independent variable (IV) and dependent variable (DV), (2) placed with all ATCB items together but after the IV and DV and (3) the seven ATCB items were ungrouped and intermittently placed throughout the survey mixed in with the IV and DV items.

> Hypothesis 6: The location of the ATCB items in a survey will have no effect on the ability of ATCB to detect CMV in an unusually strong correlation in same-source data.

Method

Participants and Procedure. The relationship between proactive personality and voice was chosen as the framework in which to investigate the placement of ATCB as a marker variable. As before, the variables are perceptual, the scales have some degree of social desirability, and most items are positively worded. A benchmark of sorts can be found in Fuller and Marler's (2009) meta-analysis. They found differences in the magnitude of correlations when same-source data versus multi-source data were used for this relationship. In same-source data, the uncorrected correlation was 0.26 and after corrections for unreliability and sampling error, it was 0.31. As a point of reference, these correlations in multi-source data were 0.11 and 0.13, respectively.

Amazon.com's MTurk was used to recruit 642 participants. The survey description was "Answer a few questions about your work attitudes," and only U.S.-based MTurkers with a 95% or greater approval rating were pre-qualified to take the survey. MTurk workers who completed a prior study for this manuscript were prohibited from participating through the use of MTurk qualification requirements, and all responses were anonymous, cross-sectional, and self-reported. The survey was hosted on Qualtrics software. Upon agreeing to the consent form, participants were randomly assigned within Qualtrics to one of the three aforementioned experimental conditions. Respondents were not allowed to take the current survey more than once.

Of the 642 respondents, the mean age was 36.43 with a range from 18 to 71 years old. The sample was 65% male and the self-reported race or ethnicity was 78.7% White, 10.9% Black, 4.5% Hispanic, 3.1% Asian, 1.7% American Indian, and 1.2% other. All respondents were currently employed in the US, with 88.6% being full-time workers. The distribution of highest achieved educational level was as follows: 0.2% less than high school diploma, 6.7% with high school diploma, 7.6%% with some college, 4.8% with a two-year college degree, 59.2% with a four-year degree, 19.6% with a professional degree, and 1.9% with a doctorate. In this sample, 94.2% completed the survey on a computer, 0.6% on a tablet, and 5.1% on a smart phone. Thirty-three participants failed to respond to one or more non-demographic survey item reducing the total sample size to 609 via listwise deletion. Final sample sizes for the three conditions were 192 for treatment condition one (ATCB items placed as a group between the IV group of items and the DV group of items), 223 for condition two (ATCB as a group placed after both the IV and DV groups), and 194 for condition three (ATCB items mixed in with IV and DV items).

Measures. The placement of the ATCB varied based on experimental condition. The items that measured demographic characteristics (age, race, sex, education, and job title) were at the end of the survey in every treatment condition. The response scale for all substantive non-demographic items was a 7-point Likert scale with verbal anchors on every point ranging from 1 = strongly disagree to 7 = strongly agree.

Proactive Personality. The ten-item version of Bateman and Crant's (1993) scale as validated by Seibert et al. (2001) was used to measure proactive personality. A sample item is "If I see something I don't like, I fix it." Coefficient alpha reliability for these scores was 0.882.

Voice. Voice was measured using Van Dyne and LePine's (1998) six-item scale, adjusted for self-reporting (e.g., "I develop and make recommendations concerning issues that affect my workgroup"). Coefficient alpha for the scores was 0.855.

Attitude Toward the Color Blue. The previously developed seven-item ATCB scale in Appendix A was used here. Alpha for these scores was 0.913.

Analysis. These data were univariate normal, with item skewness for proactive personality, voice, and ATCB ranging from −1.04 to −0.84, −1.08 to −0.88, and −1.45 to −0.99, respectively. Kurtosis values for items on these three scales ranged from 0.25 to 0.61, 0.08 to 0.97, and 0.75 to 2.33, respectively. However, the scaling correction factor for the three treatment conditions was 1.56 for the first, 1.52 for the second, and 1.51 for the third condition. For multivariate normal data the correction factor should be 1.0. Therefore, the Satorra-Bentler (S-B) corrections were implemented. As in Study Four, model fit comparisons required adjustments because of the S-B corrections (Satorra, 2000).

The same steps followed in Study Four above were followed in this study. Additionally, 95% confidence intervals around the focal correlation between proactive personality and voice were used to assess the statistical significance of any differences in that correlation in the three treatment conditions.

Results

Marker Technique Results. The basic CFA model allowed all three constructs to freely covary, forced each item onto its intended construct, and set error variances to be uncorrelated. For each latent construct, all items were freely estimated so the metric was set by constraining factor variances to one. In each treatment condition, the supplemental fit indices were quite good with the CFI ranging from 0.914 to 0.933, the SRMR ranged from 0.055 to 0.068, and the RMSEA values ranged from 0.044 to 0.055. The latent correlation between proactive personality and voice was 0.812 (95% C.I.: 0.724, 0.899) in condition one with grouped ATCB items placed between the grouped proactive personality items and the grouped voice items. In condition two, this focal correlation was 0.797 (C.I.: 0.627, 0.967) which had the grouped ATCB items placed after the grouped proactive personality items and the grouped voice items. In the third condition, the correlation was 0.845 (C.I.: 0.770, 920) in which ATCB items were placed within and between items measuring proactive personality and voice. Each correlation point estimate was within the confidence intervals of the other point estimates suggesting that there were no statistically significant differences at alpha = 0.05 between the three conditions. In fact, for each Method (i.e., Baseline, Method-C, Method-U, Method-R) model the point estimates of every correlation was firmly within the confidence intervals of every other point estimate for every Method model in all three conditions suggesting that regardless of model constraints or experimental condition, the focal correlations were not statistically different.

In condition one (ATCB inserted between the two substantive variables), the best fitting model was Method-C which constrained the paths from the marker variable to each substantive item to equivalency. When freeing those paths in the Method-U model, the change in fit was non-significant. The Method-R restricted the focal correlation between proactive personality and voice to that which was determined in the Baseline model with no significant difference resulting in comparison to the Method-C model. Therefore, in condition one, CMV was present and affected the substantive items similarly, but was not biasing.

In condition two (ATCB after substantive variables), the best fitting model was the Method-U model which allowed cross-loadings from the marker variable to the substantive items to be freely estimated. As in condition one, the imposition of a fixed value for the focal latent construct correlation had no effect on model fit (Method-R versus Method-C). Thus, in condition two, CMV was present but did affect the substantive items differently and was therefore biasing.

In condition three (ATCB intermixed with substantive variables), the best fitting model was again the Method-U model and as in the other two treatment conditions of the experiment the fixing of the focal correlation in the Method-R model had no effect on model fit. Therefore, in condition three, CMV was present and did affect the substantive items differently and was therefore biasing. With these results in mind, the placement of the ATCB items in different locations in the survey had no effect on the ability of the ATCB marker variable to detect CMV providing support for Hypothesis 6. See Table 6 for all of the fit statistics of the models in the experiment.

Reliability Decomposition Results. In the next phase of analysis, the reliability of the constructs was decomposed. In treatment condition one, the best fitting model was Method-C which yielded a total reliability for proactive personality, voice, and ATCB of 0.875, 0.870, and 0.913, respectively. For proactive personality, the decomposed substantive and method reliabilities were 0.678 and 0.197, respectively. For voice, those values were 0.725 and 0.145, respectively. Overall, the percentage of reliability attributable to the marker variable was 22.55% for proactive personality and 16.69% for voice.

In treatment condition two, the best fitting model was Method-U which yielded a total reliability for proactive personality, voice, and ATCB of 0.875, 0.850, and 0.894, respectively. For proactive personality, the decomposed substantive and method reliabilities were 0.724 and 0.151, respectively.

Table 6. Study Five Fit Indices and Model Comparison Tests Using the CFA Marker Technique in Three Treatment Conditions.

	S-B χ^2	*df*	$\Delta\chi^2$	Δdf	CFI	SRMR	RMSEA (90% C.I.)	r_{xy} (95% C.I.)
CFA model 1								
Condition 1 (middle)	354.424***	227	--	--	0.914	0.057	0.054 (0.043, 0.065)	0.812 (0.724, 0.899)
Condition 2	323.755***	227	--	--	0.933	0.068	0.044 (0.032, 0.054)	0.797 (0.627, 0.967)
Condition 3 (mixed)	360.229***	227	--	--	0.925	0.062	0.055 (0.044, 0.065)	0.845 (0.770, 0.920)
Baseline model 2								
Condition 1 (middle)	382.270***	243	--	--	0.906	0.162	0.055 (0.044, 0.065)	0.811 (0.724, 0.898)
Condition 2	344.568***	243	--	--	0.930	0.137	0.043 (0.032, 0.053)	0.797 (0.629, 0.965)
Condition 3 (mixed)	370.248***	243	--	--	0.929	0.132	0.052 (0.041, 0.062)	0.845 (0.772, 0.918)
Method-C model 3								
Condition 1 (middle)	353.768***	242	20.758[a]***	1[a]	0.925	0.061	0.049 (0.038, 0.060)	0.758 (0.639, 0.878)
Condition 2	308.194***	242	35.827[a]***	1[a]	0.954	0.090	0.035 (0.022, 0.046)	0.763 (0.570, 0.956)
Condition 3 (mixed)	345.734***	242	27.404[a]***	1[a]	0.942	0.067	0.047 (0.035, 0.058)	0.826 (0.743, 0.909)
Method-U model 4								
Condition 1 (middle)	342.551***	227	6.008[b] *ns*	15[b]	0.922	0.054	0.051 (0.040, 0.062)	0.756 (0.633, 0.878)
Condition 2	264.063***	227	47.522[b]***	15[b]	0.974	0.043	0.027 (0.004, 0.040)	0.760 (0.567.953)
Condition 3 (mixed)	321.238***	227	26.513[b]*	15[b]	0.947	0.048	0.046 (0.034, 0.058)	0.825 (0.743, 0.908)
Method-R model 5								
Condition 1 (middle)	342.579***	228	6.181[b] *ns*	14[b]	0.923	0.056	0.051 (0.040, 0.062)	0.811[d]
Condition 2	260.734***	228	6.831[c] *ns*	1[c]	0.977	0.044	0.025 (0.000, 0.039)	0.797[d]
Condition 3 (mixed)	321.032***	228	2.040[b] *ns*	1[b]	0.948	0.048	0.046 (0.034, 0.057)	0.845[d]
Method-S(0.05) model 6								
Condition 1 (middle)	357.178***	243	0.0[b] *ns*	1[b]	0.923	0.079	0.049 (0.038, 0.060)	0.756 (0.634, 0.878)
Condition 2	274.875***	243	9.675[c] *ns*	16[c]	0.978	0.107	0.024 (0.000, 0.038)	0.765 (0.576, 0.954)
Condition 3 (mixed)	335.690***	243	9.580[c] *ns*	16[c]	0.948	0.104	0.044 (0.032, 0.055)	0.830 (0.750, 0.909)
Method-S(0.01) model 7								
Condition 1 (middle)	359.537***	243	5.640[b]*	1[b]	0.922	0.095	0.050 (0.039, 0.061)	0.757 (0.636, 0.878)
Condition 2	280.759***	243	16.440[c] *ns*	16[c]	0.974	0.143	0.026 (0.003, 0.039)	0.771 (0.585, 0.956)
Condition 3 (mixed)	339.839***	243	16.217[c] *ns*	16[c]	0.946	0.137	0.045 (0.033, 0.056)	0.834 (0.756, 0.911)

Note: Treatment condition sample sizes $n_1 = 192$, $n_2 = 223$, $n_3 = 194$. Condition 1 (middle) = ATCB items as a group between the independent variable (IV) items and the dependent variable DV) items, Condition 2 = ATCB items as a group after all IV and DV items, Condition 3 (mixed) = ATCB items inserted within and between items measuring IV and DV. $\Delta\chi^2$ values are not simple difference tests and required special calculations (Satorra, 2000); r_{xy} is correlation between proactive personality and voice; *ns* = non-significant.

[a] vs. Baseline.

[b] vs. Method-C.

[c] vs. Method-U.

[d] Fixed to the value from the Baseline Model.

***$p < .001$.

For voice, those values were 0.722 and 0.128, respectively. Overall, the percentage of reliability attributable to the marker variable was 17.22% for proactive personality and 15.02% for voice.

In treatment condition three, the best fitting model was again Method-C which yielded a total reliability for proactive personality, voice, and ATCB of 0.904, 0.869, and 0.933, respectively. For proactive personality, the decomposed substantive and method reliabilities were 0.760 and 0.144, respectively. For voice, those values were 0.724 and 0.145, respectively. Overall, the percentage of reliability attributable to the marker variable was 15.90% for proactive personality and 16.68% for voice.

The total reliabilities were remarkably similar and strong across the three treatment conditions for the three variables ranging from 0.850 to 0.933. After decomposition the mean substantive reliability across conditions fell to a weaker level of 0.720 for proactive personality and 0.724 for voice. The mean percentage of reliability attributable to the ideal marker variable for scores on each focal instrument was 18.56% for proactive personality and 16.13% for voice. See Table 7 for these reliability statistics.

Sensitivity Analysis Results. The sensitivity analysis required a comparison of both the Method-S(0.05) and the Method-S(0.01) models to the better fitting of the Method-U or Method-C model in the three experimental conditions. In treatment condition one, the retained model was Method-C. These comparisons resulted in S-B $\Delta\chi^2 = 0.0$ ($\Delta df = 1$, *ns*) for Method-S(0.05) and S-B $\Delta\chi^2 = 5.640$ ($\Delta df = 1$, $p < .05$) for Method-S(0.01). Thus, fixing marker variable loadings to the substantive indicators to unstandardized values at the high end of the 95% intervals had no impact on the interpretation of the correlation between proactive personality and voice. Fixing the paths to the higher end (99% confidence interval) only slightly affected model fit but there was no significant change in the focal correlation. Additionally, for the Method-S(0.01) model the supplemental fit indices were all worse than in the Method-C model. Therefore, the results for fixing the paths to the high end of the 99% confidence interval are somewhat mixed.

In treatment condition two, the retained model was Method-U. These comparisons resulted in S-B $\Delta\chi^2 = 9.675$ ($\Delta df = 16$, *ns*) for Method-S(0.05) and S-B $\Delta\chi^2 = 16.440$ ($\Delta df = 16$, *ns*) for Method-S(0.01). For both Method-S models fixing marker variable loadings to the substantive indicators to unstandardized values at the high end of the 95% and 99% confidence intervals had no impact on the interpretation of the correlation between proactive personality and voice which ranged from 0.760 to 0.771 for all three models.

In treatment condition three, the retained model was Method-U. These comparisons resulted in S-B $\Delta\chi^2 = 9.580$ ($\Delta df = 16$, *ns*) for Method-S(0.05) and S-B $\Delta\chi^2 = 16.217$ ($\Delta df = 16$, *ns*) for Method-S(0.01). Thus, fixing marker variable loadings to the substantive indicators to unstandardized values at the high end of the 95% and 99% confidence intervals had no impact on the interpretation of the correlation between proactive personality and voice which ranged from 0.825 to 0.834 in the three models.

Discussion

These findings bolster those from Study 4 in indicating that ATCB can detect CMV in same source data. Results also indicate that placement of ATCB does not produce a meaningful difference in the detection of CMV but does influence identification of bias. The conclusion from this finding is that researchers have some initial evidence that any location of the ATCB scale in a survey is acceptable. Further, this finding points to the need for more research on the notion that inclusion of a marker variable like ATCB between independent and dependent variables can create a psychological separation. This procedural remedy recommended by Podsakoff et al. (2003) has not been directly studied, and findings here indicate that it may not be effective.

Table 7. Studies Four and Five Reliability Decomposition.

Latent variable	Total reliability	Decomposed reliability: Substantive reliability	Method reliability	% Reliability marker variable
Study four				
Organizational commitment	0.945	0.836	0.109	11.53%
Job satisfaction	0.793	0.704	0.088	11.10%
Attitude toward the color blue	0.936	0.936	--	--
Study five				
Proactive personality (condition 1)	0.875	0.678	0.197	22.55%
Proactive personality (condition 2)	0.875	0.724	0.151	17.22%
Proactive personality (condition 3)	0.904	0.760	0.144	15.90%
Voice (condition 1)	0.870	0.725	0.145	16.69%
Voice (condition 2)	0.850	0.722	0.128	15.02%
Voice (condition 3)	0.869	0.724	0.145	16.68%
Attitude toward the color blue (condition 1)	0.913	0.913	--	--
Attitude toward the color blue (condition 2)	0.893	0.893	--	--
Attitude toward the color blue (condition 3)	0.933	0.933	--	--

Note: Condition 1 placed the grouped ATCB items between the grouped independent variable (IV) items and the grouped dependent variable (DV) items. Condition 2 placed the grouped ATCB items after all of the grouped IV and DV items. Condition 3 placed the ungrouped ATCB items intermittently throughout the survey mixed in with the IV and DV items.

General Discussion

As researchers are increasingly mindful of the degree to which the CMV may bias results, the current study provides robust evidence of the reliability, validity, and usefulness of the newly developed scale measuring attitude toward the color blue. In a series of studies that follow best practices for scale development (Hinkin, 1998), ATCB was established as an ideal marker variable that likely can be used in a wide variety of social science studies. Unlike non-ideal marker variables that have been identified in research (Simmering et al., 2015), ATCB is perceptual, is measured on a Likert scale, and is highly unlikely to be theoretically related to most social science variables.

Research may still benefit from procedural steps to avoid CMV, such as collecting multi-source data or introducing a time lag between predictor and criterion measures (Podsakoff et al., 2003). However, such approaches should not automatically be considered a superior way to address method effects. Procedural remedies may introduce other method variance errors (Spector et al., 2019), and there are there are circumstance in which same-source cross-sectional data is most appropriate to research questions (Chan, 2009). Thus, in these cases, an efficacious post hoc test such as the CFA Marker Technique can be a useful tool to identify CMV in data. Yet, a review of research using marker variables indicates that the choice of marker variable is both difficult and often poorly managed (Simmering et al., 2015). Thus, the primary contribution of the current study is the development of an ideal marker variable scale that can be used in almost any social science research.

One commonly used approach to identifying CMV in data is the use of variables that are presumed to capture CMV, such as social desirability and positive affectivity (Simmering et al., 2015). Despite these variables capturing some types of response bias, they have limitations. Notably, depending on the variables in the study, a researcher risks capturing substantive variance (Spector, 2006). In the focal relationship between organizational commitment and job satisfaction examined in Study Four, positive affectivity would likely be meaningfully related to both variables and controlling for positive affectivity would likely reduce the true variance captured in that relationship. This indicates the strength of the use of a marker variable such as ATCB, which would not capture substantive variance in this relationship.

One procedural approach to reducing CMV recommended by Podsakoff et al. (2003) is to create psychological separation of items on a single survey for which all data are self-reported, such as placing a marker variable between scales measuring IVs and DVs. Those authors point to three benefits of such separation: (1) that it "… should reduce biases in the retrieval stage of the response process by eliminating the saliency of any contextually provided retrieval cues" (p. 888), (2) that respondents' abilities to remember responses to prior items when completing later items is reduced, and (3) that prior responses to survey items become less salient when answering later items. Thus, the inclusion of a variable such as ATCB on a single self-reported survey may influence a respondent to "switch gears" and therefore interrupt response patterns, particularly because it is unusual to include items regarding color on a social science survey. We tangentially tested this notion in Study Five with different ATCB placement conditions, but preliminary results indicate that this type of psychological separation within a single survey may not produce meaningful differences in responses. Notably, the type of CMV produced (congeneric versus noncongeneric) seemed to differ in our study conditions, but the overall influence of CMV was not markedly different.

The findings of this series of studies allows for recommendations for researchers and reviewers. First, both parties are encouraged to consider specific possible sources of CMV relevant to the study, rather than the specter of omnibus CMV deriving from same source responses (Spector et al., 2019). When the presence of common method effects is likely to bias findings, the CFA Marker Technique applied with an ideal marker variable is likely to detect it (Richardson et al., 2009; Williams et al., 2010). However, because there are other sources of both common and uncommon method variance, researchers should not see inclusion and analysis of a marker variable as a panacea for all CMV. Thus, researchers are encouraged to consider the effects of items, the survey, and the measurement context and take steps to design surveys with procedural efforts that are both likely to mitigate CMV and to improve survey validity (Podsakoff et al., 2003). Researchers are encouraged to take an even broader view of the issue of CMV and recognize that there is a larger debate surrounding the nature and likelihood of CMV, including the influence of uncommon method variance (Spector et al., 2019).

A second recommendation is that researchers who use the CFA Marker Technique choose an ideal marker variable and include it in a survey with same source data *a priori*. As noted in Simmering et al. (2015), use of the CFA Marker Technique grew over time, yet many of the marker variables used in published research did not meet the criteria of the ideal marker. Thus, identification of a psychometrically sound marker that is unlikely to be theoretically related to most social science variables increases the opportunity for researchers to make wise choices about marker variables. Although we acknowledge that there may be a very small number of studies in which ATCB should not be used (e.g., in a sample of employes that work for an organization that identifies strongly with the color blue, such as IBM, sometimes called "Big Blue"), this constraint is likely to be rare. The current study provides evidence that ATCB can be successfully used as an ideal marker variable for a wide variety of workplace investigations, as it has demonstrated reliability, construct validity, and discriminant validity. Notably, our results indicate that the placement of the marker variable does not seem to matter in its ability to detect CMV.

In using the ATCB scale, if items are presented in a block, these instructions should be used: "Please indicate your level of agreement or disagreement with the following statements by using the response scale below." An absence of information about the scale precludes priming respondents, which is likely to capture less crystallized responses about the specific color and rather be more likely to measure common rater effects. Researchers should use a Likert scale to measure responses, with anchors indicating level of agreement (i.e., strongly disagree to strongly agree). The choice of number of scale points depends on researcher preference, noting that presenting different scale formats may reduce CMV (Podsakoff et al., 2003). As is evidenced in Study Five, there is initial evidence that placement of a marker variable in a survey is not meaningful; yet, future research may address this question more fully.

Finally, researchers and reviewers are encouraged to familiarize themselves with the CFA Marker Technique. To that end, we have provided the Mplus code for the series of tests in this approach as a means to better understand the use of it. Our review of articles published in 2020 that cited Simmering et al. (2015) and/or Williams et al. (2010) identified a handful of articles that included a marker variable, but did not use it in the CFA Marker Technique, instead simply examining correlations among the marker and substantive variables or including it as a control variable. The efficacy of these other uses of marker variables is unknown, and researchers are encouraged to apply the full CFA Marker Technique.

There are several limitations in this research. First, some may question the use of survey panels such as Amazon.com's MTurk in regards to both the data quality and generalizability of findings. As can be seen, data collected from these participants yielded very high correlations. Yet, investigations into this relatively new subject pool indicate that data tends to be as reliable as that of other popular sources (i.e., college students) and that Mechanical Turk participants tend to be more representative of the U.S. population than most in-person convenience samples (Behrend et al., 2011; Berinsky et al., 2012). Additionally, we employed best practices for data collection in MTurk (Aguinis et al., 2021). A limitation regarding the use of marker variables in general is that, despite our anticipation that ATCB is capturing a number of common rater effects that may be present in data, however many of these are unmeasurable and it is therefore not possible to know which are addressed with a marker variable.

Future research on ATCB and marker variables in general is warranted to address remaining questions. First, as discussed above, there is preliminary evidence that placement of a marker variable is not meaningful in terms of CMV detection but was different in terms of identifying bias. This finding requires more research, particularly as there have been scant investigations of the efficacy of procedural approaches to avoiding CMV. Further, research findings regarding item positioning on electronically-delivered surveys indicates that responses may be more positive to items that appear in a higher position on a screen (Tourangeau et al., 2013). Thus, responses to marker variables within a survey may differ based on whether the survey is completed on paper, computer, or a smartphone, which has not been empirically investigated.

A fruitful area for future research is how insufficient effort responding (IER) may relate to CMV. IER occurs when "the respondent answers a survey measure with low or little motivation to comply with survey instructions, correctly interpret item content, and provide accurate responses," (Huang et al., 2012, p. 100). This rater behavior could be related to the common rater effects that influence CMV and are also captured by ATCB. Yet, to our knowledge, no published article has explicitly conceptualized how one can simultaneously capture CMV and IER. Lastly, the ATCB scale may be useful for differentiating CMV from uncommon method variance, expanding on the work of Spector et al. (2019).

In conclusion, the current study is the first to develop and validate an ideal marker variable for use in statistical detection of CMV in same-source data. Attitude Toward the Color Blue exhibits reliability, validity, and is distinct from other measures. Evidence indicates that it can detect CMV in strongly correlated same-source correlations. And, unlike other marker variables, ATCB is likely to be theoretically unrelated to a wide variety of social science variables. Researchers who are investigating constructs unrelated to the color blue can have confidence that the ATCB ideal marker variable can effectively identify common method variance in data.

Appendix

Attitude Toward the Color Blue

1. Blue is a beautiful color.
2. Blue is a lovely color.
3. Blue is a pleasant color.
4. The color blue is wonderful.
5. Blue is a nice color.
6. I think blue is a pretty color.
7. I like the color blue.

Author Note

A previous version of this manuscript was accepted for presentation at the Academy of Management conference in 2020.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Brian K. Miller https://orcid.org/0000-0001-6294-5235

Notes

1. Mplus code can be found at https://doi.org/10.18738/T8/97MGED.
2. The NOSQ, to our knowledge, has not been used as a marker variable, despite it having many of the characteristics that researchers would seek in one. A review of the articles citing those that propose the NOSQ (Eschleman & Bowling, 2011; Johnson et al., 2011; and Weitz, 1952) indicate that the NOSQ has been used to control for trait affectivity to reduce CMV one time, but in no instance did any of the articles in the citation search indicate that it had been used as a marker variable. In contrast, the original ATCB scale, which appeared in a conference presentation by Miller and Chiodo in 2008 has been used as a marker variable in at least 15 articles.

References

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management*, *47*(4), 823–837. https://doi.org/10.1177/0149206320969787

Ashkanasy, N. M. (2008). Submitting your manuscript. *Journal of Organizational Behavior*, *29*(3), 263–264. https://doi.org/10.1002/job.524

Bateman, T. S., & Crant, J. M. (1993). The proactive component of organizational behavior. *Journal of Organizational Behavior*, *14*(2), 103–118. https://doi.org/10.1002/job.4030140202

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*(3), 800–813. https://doi.org/10.3758/s13428-011-0081-0

Bem, S. L. (1974). The measure of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155–162. https://doi.org/10.1037/h0036215

Bentler, P. M. (1998, March 10). *Re: Kurtosis, residuals, fit indices*. [Message posted to SEMNET listserv]. Msg 011264. http://bama.ua.edu/cgibin/wa?A2=ind9803&L=semnet&P=R10144&I=1.

Bentler, P. M., & Wu, E. J. (2002). *EQS for Windows user's guide*. Multivariate Software.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Castille, C., Crawford, W., & Simmering, M. J. (2017). *You gotta keep 'em separated: The efficacy of proximal remedies for method variance*. Paper Presented at the Southern Management Association Meeting, St. Pete Beach, FL.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance, & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 309–336). Routledge, Taylor, and Francis.

Chang, S. J., Van Witteloostuijn, A., & Eden, L. (2010). From the editors: Common method variance in international business research. *Journal of International Business Studies*, *41*, 178–184. https://doi.org/10.1057/jibs.2009.88

Chin, W. W., Thatcher, J. B., & Wright, R. T. (2012). Assessing common method bias: problems with the ULMC technique. *MIS Quarterly*, *36*(3), 1003–1019. https://doi.org/10.2307/41703491

Comrey, A. L. (1973). *A first course in factor analysis*. Academic Press.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Dalal, D. K., & Carter, N. T. (2015). Negatively worded items negatively impact survey research. In C. E. Lance, & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 112–123). Routledge, Taylor, and Francis.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*(3), 292–307. https://doi.org/10.1037/1082-989x.2.3.292

Eschleman, K. J., & Bowling, N. A. (2011). A construct validation of the neutral objects satisfaction questionnaire (NOSQ). *Journal of Business and Psychology*, *26*(4), 501–515. https://doi.org/10.1007/s10869-010-9206-1

Fuller, B., Jr., & Marler, L. E. (2009). Change driven by nature: A meta-analytic review of the proactive personality literature. *Journal of Vocational Behavior*, *75*(3), 329–345. http://dx.doi.org/10.1016/j.jvb.2009.05.008

Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common methods variance detection in business research. *Journal of Business Research*, *69*(8), 3192–3198. https://doi.org/10.1016/j.jbusres.2015.12.008

Fuller, J. B., Patterson, C. E., Hester, K. I. M., & Stringer, D. Y. (1996). A quantitative review of research on charismatic leadership. *Psychological reports*, *78*(1), 271–287. https://doi.org/10.2466/pr0.1996.78.1.271

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Erlbaum.

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology*, *60*(2), 159. https://doi.org/10.1037/h0076546

Hair, J. E., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Prentice Hall.

Henson, R. K. (1999). Multivariate normality? What is it and how is it assessed. In B. Thompson (Ed.), *Advances in social science methodology* (pp. 193–211). JAI Press.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121. https://doi.org/10.1177/109442819800100106

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, *27*, 99–114. https://doi.org/10.1007/s10869-011-9231-8

Johnson, R. E., Rosen, C. C., & Djurdjevic, E. (2011). Assessing the impact of common method variance on higher order multidimensional constructs. *Journal of Applied Psychology*, *96*(4), 744–761. https://doi.org/10.1037/a0021504

Jöreskog, K., & Sörbom, D. (2006). *LISREL 8.80* [Computer Software]. Scientific Software International.

Judge, T.A., & Bretz, R.D. (1993). Report on an alternative measure of affective disposition. *Educational and Psychological Measurement, 53*, 1095–1104. https://doi.org/10.1177/0013164493053004022

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.

Lance, C. E., Dawson, B., Birkelbach, D., & Hoffman, B. J. (2010). Method effects, measurement error, and substantive conclusions. *Organizational Research Methods*, *13*(3), 435–455. https://doi.org/10.1177/1094428109352528

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross sectional research designs. *Journal of Applied Psychology*, *86*(1), 114–121. https://doi.org/10.1037/0021-9010.86.1.114

Lovett, M., Bajaba, S., Lovett, M., & Simmering, M. J. (2018). Data quality from crowdsourced surveys: A mixed method inquiry into perceptions of Amazon's Mechanical Turk Masters. *Applied Psychology: An International Review*, *67*(2), 339–366. https://doi.org/10.1111/apps.12124

Mathieu, J. E., & Zajac, D. M. (1990). A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychological Bulletin*, *108*(2), 171. https://doi.org/10.1037/0033-2909.108.2.171

Miller, B. K., & Chiodo, B. (2008). *Academic entitlement: Adapting the equity preference questionnaire for a university setting*. Paper presented at the Southern Management Association Meeting, St. Pete Beach, FL.

Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior*, *14*(2), 224–247. https://doi.org/10.1016/0001-8791(79)90072-1

Muthén, L. K., & Muthén, B. (2018). *Mplus. The comprehensive modelling program for applied researchers: User's guide*, 5. Muthén and Muthén.

Osakwe, C. N., Palamidovska-Sterjadovska, N., Mihajlov, M., & Ciunova-Shuleska, A. (2020). Brand orientation, brand-building behavior and brand identity in SMEs: An empirical evaluation. *Marketing Intelligence & Planning*, *38*(7), 813–828. https://doi.org/10.1108/MIP-07-2019-0370

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*(3), 598–609. https://doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L. (1988). *Manual for the balanced inventory of desirable responding (BIDR-6)*. University of British Columbia. Unpublished manual.

Podsakoff, M. P., & Organ, W. D. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531–544. https://doi.org/10.1177/014920638601200408

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Randrianasolo, A., Semenov, A., Arnold, M., & Reynolds, K. (2020). A model of cultural intelligence and global identity. *Journal of Consumer Marketing*, *37*(7), 821–832. https://doi.org/10.1108/JCM-05-2019-3244

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*, *12*(4), 762–800. https://doi.org/10.1177/1094428109332834

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis. A Festschrift for Heinz Neudecker* (pp. 233–247). Kluwer Academic Publishers.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. https://doi.org/10.1007/bf02296192

Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser, & N. Schwarz (Eds.). *Intrapersonal processes* (pp. 436–457). Blackwell.

Seibert, S. E., Kraimer, M. L., & Crant, J. M. (2001). What do proactive people do? A longitudinal model linking proactive personality and career success. *Personnel Psychology*, *54*(4), 845–874. https://doi.org/10.1111/j.1744-6570.2001.tb00234.x

Simmering, M. J., Fuller, C. M., Richardson, H. A., Ocal, Y., & Atinc, G. M. (2015). Marker variable choice, reporting, and interpretation in the detection of common method variance: A review and demonstration. *Organizational Research Methods*, *18*(3), 473–511. https://doi.org/10.1177/1094428114560023

Spector, P. E. (1987). Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology*, *72*(3), 438–443. https://doi.org/10.1037/0021-9010.72.3.438

Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior*, *15*(5), 385–392. https://doi.org/10.1002/job.4030150503

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods*, *9*(2), 221–232. https://doi.org/10.1177/1094428105284955

Spector, P. E., & Brannick, M. T. (1995). The nature and effects of method variance in organizational research. In C. L. Cooper, & I. T. Robertson (Eds.). *International review of industrial and organizational psychology* (pp. 249–274). John Wiley.

Spector, P. E., Rosen, C. C., Richardson, H. A., Williams, L. J., & Johnson, R. E. (2019). A new perspective on method variance: A measure-centric approach. *Journal of Management*, *45*(3), 855–880. https://doi.org/10.1177/0149206316687295

Tabachnik, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Pearson.

Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. *Personnel Psychology*, *46*(2), 259–293. https://doi.org/10.1111/j.1744-6570.1993.tb00874.x

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.

Thompson, B., & Daniel, L. (1996). Factor analytic evidence for the construct validity of scores: An historical overview and some guidelines. *Educational and Psychological Measurement*, *56*(2), 197–208. https://doi.org/10.1177/0013164496056002001

Thurstone, L. L. (1935). *The vectors of the mind*. University of Chicago Press.

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2013). "Up means good" the effect of screen position on evaluative ratings in web surveys. *Public Opinion Quarterly*, *77*(S1), 69–88. https://doi.org/10.1093/poq/nfs063

Tourangeau, R., Rips, L.J., & K. Rasinski (2000). *The psychology of survey response*. Cambridge University Press.

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, *37*(1): 1–2. https://doi.org/10.1080/01973533.2015.1012991

Vandenberg, R. J. (2006). Introduction: Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? *Organizational Research Methods*, *9*(2), 194–201. https://doi.org/10.1177/1094428105285506

Van Dyne, L., & LePine, J. A. (1998). Helping and voice extra-role behavior: Evidence of construct and predictive validity. *Academy of Management Journal*, *41*(1), 108–119. https://doi.org/10.5465/256902

Wall, A. E., Fuller, C. M., & Simmering, M. J. (2015). *Manipulating and measuring common method variance in same source data: An experiment*. Paper Presented at the Academy of Management Meeting, Philadelphia, PA.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1107. https://doi.org/10.1037/0022-3514.54.6.1063

Weitz, J. (1952). A neglected concept in the study of job satisfaction. *Personnel Psychology*, *5*(3), 201–205. https://doi.org/10.1111/j.1744-6570.1952.tb01012.x

Werts, C. E., Linn, R. L., & Joreskog, K. G. (1974). Interclass reliability estimates: Testing structural assumptions. *Educational and Psychological Measurement*, *34*(1), 25–33. https://doi.org/10.1177/001316447403400104

West, S. G., Finch, J. D. F., & Curran, P. J. (1995). Structural equation models with nonnormal data. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 56–75). Sage.

Williams, L. J., & Anderson, S. E. (1994). An alternative approach to method effects by using latent-variable models: Applications in organizational behavior research. *Journal of Applied Psychology*, *79*(3), 323–331. https://doi.org/10.1037/0021-9010.79.3.323

Williams, L. J., & Brown, B. K. (1994). Method variance in organizational behavior and human resources research: Effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes*, *57*(2), 185–209. https://doi.org/10.1006/obhd.1994.1011

Williams, L. J., Cote, J. A., & Buckley, M. R. (1989). Lack of method variance in self-reported affect and perceptions at work: Reality or artifact? *Journal of Applied Psychology*, *74*(3), 462–468. https://doi.org/10.1037/0021-9010.74.3.462

Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method variance and marker variables: A review and comprehensive CFA marker technique. *Organizational Research Methods*, *13*, 477–514. https://doi.org/10.1177/1094428110366036

Williams, L. J., & McGonagle, A. K. (2016). Four research designs and a comprehensive analysis strategy for investigating common method variance with self-report measures using latent variables. *Journal of Business and Psychology*, *31*(3), 339–359. https://doi.org/10.1007/s10869-015-9422-9

Williams, L. J., & O'Boyle, E. H. (2015). Ideal, nonideal, and no-marker variables: The confirmatory factor analysis (CFA) marker technique works when it matters. *Journal of Applied Psychology*, *100*(5), 1579. https://doi.org/10.1037/a0038855

Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin, & A. Tesser (Eds.). *The construction of social judgments* (pp. 37–65). Erlbaum.

Author Biographies

Brian K. Miller is a professor of Management at Texas State University who studies measurement issues associated with maladaptive personality traits. He earned a PhD from the University of Houston in 2002 and a post-doctoral Masters degree in Quantitative Methods from the University of Texas at Austin in 2016. He is the former associate editor for Methodology at the *Journal of Managerial Psychology and at Group and Organization Management*.

Marcia J. Simmering is the Francis R. Mangham Endowed professor of Management at Louisiana Tech University. Her research primarily addresses research methods (specifically common method variance and control variables) and human resources. Her work has appeared in *Organizational Research Methods*, the *Academy of Management Journal*, and the *Journal of Applied Psychology*. She has twice won the best paper award for the Southern Management Association conference.