JPART 25:285–308

How (Not) to Solve the Problem: An Evaluation of Scholarly Responses to Common Source Bias

Nathan Favero, Justin B. Bullock
Texas A&M University

ABSTRACT

Public administration scholars are beginning to pay more attention to the problem of common source bias, but little is known about the approaches that applied researchers are adopting as they attempt to confront the issue in their own research. In this essay, we consider the various responses taken by the authors of six articles in this journal. We draw attention to important nuances of the common measurement issue that have previously received little attention and run a set of empirical analyses in order to test the effectiveness of several proposed solutions to the common-source-bias problem. Our results indicate that none of the statistical remedies being used by public administration scholars appear to be reliable methods of countering the problem. Currently, it appears as though the only reliable solution is to find independent sources of data when perceptual survey measures are employed.

Public administration scholars make extensive use of surveys because of their flexibility in allowing researchers to measure many important variables at a relatively low cost. As with any tool, however, the survey method does have its limitations. Research in the fields of psychology and business (Baumgartner and Steenkamp 2001; Campbell and Fiske 1959; Cote and Buckley 1987; Doty and Glick 1998; Podsakoff et al. 2003; Podsakoff, MacKenzie, and Podsakoff 2012; Podsakoff and Organ 1986; Williams, Cote, and Buckley 1989) and more recently in public administration (Meier and O'Toole 2013b) has explored a particular limitation of the survey method: common method bias. Common method bias is a biasing of results (which could be in the form of false positives from hypothesis tests) that is caused by two variables exhibiting related measurement error owing to a common method, such as a single survey. Although there are multiple types of common method bias, we focus in this essay on bias arising from the use of a common source (or related sources) to measure both the independent and dependent variables.

Meier and O'Toole (2013b) provided the first study devoted to the issue of common source bias in the field of public administration. The publication of their

Address correspondence to the author at nathan.favero@tamu.edu.

doi:10.1093/jopart/muu020
Advance Access publication May 24, 2014

study has begun to generate a discussion on the topic among applied researchers as they grapple with how their own results might be affected by common source bias. In some cases, scholars have adopted unique methodological approaches aimed at detecting or overcoming common source bias, but the effectiveness of such methods is generally not well understood (Podsakoff et al. 2003; Podsakoff, MacKenzie, and Podsakoff 2012). Other public administration scholars have attempted to skirt the problem by arguing that a common source of measurement should not be problematic for their particular research question. Still others continue to completely ignore the potential issues posed by utilizing a common source to measure independent and dependent variables. In this essay, we evaluate these recent responses to the common-source-bias problem by considering six articles published in the *Journal of Public Administration Research and Theory*. Our essay provides the field's first systematic evaluation of various proposed methodological remedies for potential common source issues.

We begin by offering a theoretical treatment of common source bias and why it should be a serious concern for researchers in many cases. We then take this knowledge and discuss how six articles address (or fail to address) the common-source-bias problem and consider the effectiveness of their respective approaches. As a part of this discussion, we independently test the usefulness of several of their approaches using two data sets: Texas school superintendent surveys matched with Texas Education Agency data and New York City school teacher surveys matched with administrative city and state records. It is our hope that this essay may serve as a guide to public administration scholars for not only how to evaluate the potential for common source bias to contaminate their results but also what can (or cannot) be done to effectively remedy potential problems.

THE COMMON-SOURCE-BIAS PROBLEM

Common source bias can easily be explained with the aid of a few simple equations which will allow us to make some important observations about when bias is or is not likely to exist given the measures being used. Public administration scholars often rely on survey responses to measure behavior or organizational characteristics. Many of these measures are perceptual, meaning that they require respondents to make either a subjective judgment or an estimation. Perceptions imply that two different respondents might describe the same phenomenon or characteristic differently. For some researchers, these differences in perceptions may themselves be the subject of study (which we discuss further below), but in many public administration applications, perceptions simply provide a convenient means of measuring variation in organizations or behavior. Such perceptual measures will consist of two parts: the actual value of whatever variable the researcher wishes to measure (X) plus some perceptual error (p_error_X):

$$perception(X) = X + p\_error_X \tag{1}$$

When perceptions are highly accurate, the value of the error term (p_error_X) will tend to be very small. Existing research on both public and private managers has shown that perceptions of the performance of one's own organization appear to exhibit rather large errors (see Meier and O'Toole 2013a).

Common method bias arises when estimating the relationship between two variables that have correlated measurement errors. We will illustrate the problem of common method bias by demonstrating the effect that correlated measurement errors can have on the covariance of two perceptual measures. Mathematically, it can be shown that the covariance of two perceptual measures will reflect not only the covariance between the two true underlying variables but also the covariance between the perceptual errors of the two variables. Assuming the perceptual errors are independent from the true values of X and Y:

$$COV\left(perception\left(X\right), perception\left(Y\right)\right) = COV\left(X,Y\right) + COV\left(p\_error_X,\ p\_error_Y\right) \quad (2)$$

If the two sets of perceptual errors are correlated ($COV\left(p\_error_X, p\_error_Y\right) \neq 0$), use of the perceptual variables will produce biased estimates of the covariance of X and Y. Positively correlated errors will yield positively biased estimates. A positive bias may have the effect of either attenuating or inflating the covariance estimate depending on the direction of the true covariance. Notice that when the true covariance between the underlying variables X and Y is equal to zero, the covariance of the perceptual measures is exactly equal to the covariance of the errors. Thus, correlated measurement errors can produce the appearance of a relationship where none exists.

As one moves beyond covariance estimates and considers estimates of correlation or regression coefficients, determining whether correlated measurement errors will attenuate, inflate, or have no effect on estimates becomes more complicated (Cote and Buckley 1988; Siemsen, Roth, and Oliveira 2010). In particular, one must account for the attenuation bias that generally results from any nonsystematic variation in the measurement errors. Lance et al. (2010) argue that this attenuation bias often counteracts the inflationary effect of common source variance, leading the authors to conclude that common source concerns are overstated. However, it is important to recognize that no attenuating effect can exist when there is no actual relationship between two variables. Public management scholars are usually interested in conducting hypothesis tests with a null hypothesis of no effect. If the null of no effect is true, common method variance can only bias estimates in an inflationary manner, which will tend to produce the appearance of a relationship where none exists. The potential for common method bias to produce false positives should be of central concern in public management applications.

The causes of common source bias may have important implications for potential remedies, which we will discuss in more detail below. For our purposes, it is useful to identify two broad categories of biasing effects which we will refer to as individual effects and environmental effects:

$$p\_error_{X,i,e,t} = individual\_bias_{X,i,t} + environmental\_bias_{X,e,t} + other\_error_{X.i,e,t} \quad (3)$$

This equation suggests that perceptual errors are a function of the biases of the individuals providing the perceptions ($individual\_bias_{X,i,t}$), the biases produced by the environments in which the individuals are located ($environmental\_bias_{X,e,t}$), and other unknown factors ($other\_error_{X,i,e,t}$). The basis for our concept of an individual bias derives directly from existing research which suggests that perceptual errors for measures of two distinct variables provided by the same individual on a perceptual survey are often correlated (e.g., Meier and O'Toole 2013b; Podsakoff et al. 2003). In other words, for the two variables X and Y, $p\_error_{X,i,e,t}$ and $p\_error_{Y,i,e,t}$ are likely to be correlated because $individual\_bias_{X,i,t}$ and $individual\_bias_{Y,i,t}$ are related (or even identical). Our concept of environmental bias is more novel to the literature, although it is somewhat similar to what Podsakoff et al. (2003) refer to as "measurement context effects." Whereas Podsakoff et al. and the literature they cite focuses on the context in which the survey research itself is situated, we wish to draw attention to the possibility that the organizational context in which an individual is situated may be related to how he/she processes and reports information about his/her workplace. Specifically, organizational culture and socialization processes (Schein 1992) may shape the way that individuals perceive and relay information about various dimensions of their organizations. Furthermore, the hiring process of an organization may tend to select employees with similar personalities or perspectives.

As indicated by the subscripts in equation (3), the biases produced by individuals and environments may vary depending on both the variable being measured (X) and the time at which the perception is reported (t). Whether the individual or environmental bias associated with a measure actually varies along these two dimensions may depend on the precise cause(s) of the bias. Of the various potential causes of individual bias (or "common rater effects") identified in the literature (see Podsakoff et al. 2003, table 2, 882), a transient mood state clearly varies over time while several other causes (e.g., implicit theories, social desirability, acquiescence bias) might be expected to produce biases that are relatively stable. A social desirability bias should primarily affect questions for which there is clearly a socially preferred answer while a positivity or negativity bias may produce a more uniform bias across a variety of perceptual questions. Environmental biases corresponding to those at the individual level might be evident at an organizational level if, for example, an organization experiences an event which alters the transient mood state of its employees or if an organization tends to hire individuals with positive affectivity.

Determining the specific cause(s) of the bias associated with a perceptual measure should provide insights regarding appropriate remedies. For example, if the bias is produced solely by a transient mood state, measures at sufficiently different points in times should exhibit independent errors. In practice, it is often difficult to rule out the possibility that there are multiple causes of bias affecting a given perceptual measure, so the assumption that bias is either fully independent or completely fixed across time or across survey items may be difficult to justify. Later in this essay, we empirically test whether or not remedies based on the assumption of fixed bias across time or survey items appear to adequately address common method concerns for two different public administration data sets.

Thus far, we have focused on the use of perceptual survey items to measure *behaviors* or *organizational characteristics*. However, researchers sometimes wish to measure *attitudes*, *perceptions*, or *affects*, such as job satisfaction, public service motivation, citizen satisfaction, or perceptions of management. In such cases, variation in how individuals make subjective judgments or estimations constitutes an important aspect of the variable of interest rather than merely the source of a measurement error.[1] A consideration of how common method bias may affect estimation of relationships between two perceptual variables lies beyond the realm of this essay, but we extend our analysis here to consider studies of the relationship between a perceptual and a nonperceptual variable. To this end, we offer the following equation:

$$\begin{aligned} COV\left(perception(X), perception(Y)\right) = COV\left(perception(X), Y\right) \\ + COV\left(perception(X), p\_error_Y\right) \end{aligned} \tag{4}$$

The estimation of the relationship between a perceptual variable ($perception(X)$) and a nonperceptual variable (Y) will be biased if the measurement error associated with the nonperceptual variable (p_error_Y) is correlated with the perceptual variable of interest ($perception(X)$). An example comes from Sharma, Yetton, and Crawford (2009), who find evidence that estimates of the correlation between perceived usefulness of technology (a perception) and use of technology (a behavior) are inflated when a perceptual measure (rather than a system-generated measure) of the behavior is used. Similarly, Brown and Benedict (2002, 562–63) question the validity of studies finding a positive relationship between police satisfaction and (perceptual measures of) police behavior or response time.

Given our analysis thus far, it should be conceptually clear how common source bias can affect studies relying on perceptual measures of both independent and dependent variables. Unfortunately, there are often no easy statistical solutions to the problem of common source bias, and the appropriateness of particular remedies depends on the specific cause(s) of the bias (Podsakoff et al. 2003; Podsakoff, MacKenzie, and Podsakoff 2012). If researchers only have one source available to them, it can be extremely difficult (if not impossible) to disentangle what portion of the estimated coefficient of the relationship between the dependent and independent variable owes to an actual relationship between the variables of interest and what portion of the estimate is due to a relationship between the measurement errors.

THE COMMON-SOURCE-BIAS PROBLEM IN PRACTICE

Over the years, public administration scholars have dealt with common sources of measurement in a number of ways. In this section, we identify seven approaches that public administration scholars have recently taken towards the common-source-bias problem: *ignoring the problem*, *adjusting interpretation of variables*, *Harman's single-factor test*, *Brewer's split sample method*, *marker variables*, *differencing*, and *finding an independent source of data*. These approaches will be discussed within the context

1 We do not mean to imply that survey measures of perceptual variables (constructs) do not contain measurement errors. Our point is simply that what is considered part of the measurement error under one research question may constitute part of the variable of interest under another research question. Methods (including exploratory and confirmatory factor analysis) have been developed to deal with measurement error of perceptual variables.

of six articles published in this journal. We will highlight how the findings in these articles may be affected by common method bias, examine the authors' responses to potential issues, and consider whether their approaches appear to reliably deal with the potential problems associated with using a common source.

Ignoring the Problem

Some scholars have continued to ignore the potential bias that a common source of data can introduce. The article "Internet, Trust in Government, and Citizen Compliance" by Im et al. (2012) uses a common survey source to measure both independent and dependent variables but ignores the potential bias this may produce. In their defense, existing public administration literature has not made it particularly clear whether or not common source bias should be a concern when investigating the relationship between a perceptual variable and a nonperceptual variable. The primary theoretical interest of Im et al. concerns the relationship between a perception (government trust) and two nonperceptual variables (Internet use and citizen compliance), although the authors also test for other relationships.

The authors' first model shows that individuals who report spending more time reading newspapers also report spending more time using the Internet. Although this model does not constitute the article's central findings, it is a straightforward example of the type of common source research design identified by Meier and O'Toole (2013b) since it tests for a relationship between two nonperceptual variables (which are measured with perceptual survey items). On an intuitive level, it seems likely that individuals who tend to overestimate the time they spend reading newspapers would also be likely to overestimate the time they spend on the Internet. This argument casts serious doubt on the authors' interpretation of the regression, which they claim indicates that time reading newspapers is positively correlated with time on the Internet.

The rest of the models of Im et al. (2012) use trust in government—a perceptual variable—as either a dependent variable or a key independent variable. Both specifications of the second model of Im et al. indicate that *perceived* Internet use time is negatively related to trust in government. This result is misleading if the misestimation of Internet time is related to reported trust in government. This relationship might exist if, for example, trust in government and low Internet usage are both considered socially desirable, meaning that people who are less susceptible to social desirability bias (all else being equal) will tend to report lower levels of trust in government and higher Internet usage, producing a negative bias. Thus, *actual* Internet usage may be unrelated to trust in government, even though *estimated* usage is related to reported trust. The third model of Im et al.—which shows that trust in government is positively related to perceived compliance—suffers from the same potential weakness. Just as Sharma, Yetton, and Crawford (2009) found that the same-source perceptual measures of technology use inflated correlation estimates, the results of Im et al. may be subject to an inflation bias. Their reliance on a common source of survey data for both independent and dependent variables leaves their findings vulnerable to potential biases that could invalidate their major substantive conclusions.

Adjusting Interpretation of Variables

Another approach to the common source problem is to adjust the interpretation of variables in order to acknowledge that the study only reveals information about the ways in which *perceptions* are related to one another. Lee, Chen, and Chou (2013) partially adopt this approach in their article entitled "Decreasing Tax Collectors' Perceived Social Loafing through Collaborative Behaviors of Taxpayers." The authors claim to circumvent the common-source-bias problem by focusing on "perceived social loafing" rather than actual social loafing. However, their other main variable of interest—collaborative behavior—is also measured as a perception, and the authors do not adjust their interpretation to reflect this fact. As demonstrated in equation (4) and through the research of Sharma, Yetton, and Crawford (2009), changing one variable to a perception does not eliminate the potential problem of common source bias.

Another problem with Lee, Chen, and Chou's (2013) article is that they are inconsistent in their treatment of the dependent variable as a perception. None of the literature cited in their "Antecedents of Social Loafing" subsection appears to address the question of how individuals form perceptions of social loafing, even though the authors claim to be identifying the "main types of factors that influence the perceived social loafing for an individual working in a group." When developing their hypotheses, the authors provide a brief explanation of how each of the three key independent variables (types of collaborative behavior) might affect social loafing. There is, however, no explanation of how these variables might affect perceived social loafing, although the term "perceived" is included in the formal statements of the hypotheses. In their discussion of findings, Lee, Chen, and Chou treat perceptions and behavior as interchangeable in at least one paragraph, going so far as to claim that their results suggest that "in order to alleviate the negative phenomenon of social loafing, government organizations should not only reform organizational structures and employee attitudes but also emphasize reduction of social loafing through direct citizen participation." If the authors wish to make such claims, they should acknowledge that common source bias may affect this conclusion. Scholars who wish to avoid critiques of their use of a common source by claiming they are only studying perceptions should truly limit their study to a consideration of perceptions.

Lee, Chen, and Chou (2013) also defend their empirical approach by arguing that social desirability bias (which can cause common source bias) is minimized because respondents are asked to describe the behavior of others rather than of themselves. However, social desirability is but one cause of common source bias, and Lee, Chen, and Chou's measure seems likely to be potentially subject to at least one other common cause of bias: positive affectivity. Positive affectivity refers to "the propensity of respondents to view themselves and the world around them in generally positive terms" (Podsakoff et al. 2003, 882). One could argue that the most parsimonious explanation of Lee, Chen, and Chou's (2013) results is that measures of perceived social loafing and perceived citizen collaboration are correlated because they reflect employees' broader perceptions of those around them. Individuals who tend to assume the best of intentions (positive affectivity) in others are more likely to report both low levels of social loafing and favorable citizen collaboration. Unfortunately, Lee, Chen, and Chou provide us with no means of determining whether perceived social loafing is related to actual citizen behavior or only to perceptions of citizen behavior.

Harman's Single-Factor Test

One of the simplest approaches to testing for a common source bias is to perform Harman's single-factor test. Jung (2013) performs this test in his article entitled "Organizational Goal Ambiguity and Job Satisfaction in the Public Sector." Three out of Jung's four measures of goal ambiguity are measured independently of the dependent variable, so Jung's general conclusion that goal ambiguity is negatively related to job satisfaction is not dependent on his argument that he has adequately dealt with the common source problem. However, his measures of mission comprehension goal ambiguity and of several control variables are derived from the same survey results that are used to measure the dependent variable (job satisfaction). Jung attempts to test for common source bias among these variables using Harman's single-factor test as well as Brewer's test (discussed below).

Harman's single-factor test requires that all of the variables (dependent and independent) being tested are placed in a factor analysis. Emergence of a single factor (or large amount of variance that can be explained by one factor) is taken as evidence that common source bias is present (Podsakoff and Organ 1986). Despite widespread use, Podsakoff et al. (2003) recommend against using this test because of substantial limitations. Most notably, the existence of multiple factors should not be taken as reliable evidence of the absence of common source bias. Jung reports that a single factor accounts for only 39% of the variance in the survey items, but if that 39% of common variance is due to common measurement error, there is a serious common-source-bias issue.

We test the effectiveness of Harman's single-factor test at detecting the common source measurement errors present among several variables which produced particularly strong biases in Meier and O'Toole's (2013b) study. We use the same 2009 Texas school superintendent survey data as Meier and O'Toole, but we also add survey results from 2011. We matched these results with corresponding records from the Texas Education Agency (TEA) for the 2008/09 and 2010/11 school years. Like Meier and O'Toole, we test for the effect of management variables on school performance, using two independent sets of standardized exam results as dependent variables (Texas Assessment of Knowledge and Skills [TAKS] and SAT/ACT test results). We then test whether or not the use of a common source alters these findings by using three perceptual measure of school performance (perceptions of TAKS performance, overall quality of education, and college bound performance).[2] We examine six perceptual management measures as independent variables (see Appendix 1).[3] In order to evaluate whether or not each of the six management items exhibits a relationship with each of the five dependent variables, we ran a series of 30 regressions. Each regression contained one measure of

2 Although the perceptual questions ask about relative (rather than absolute) levels of performance, Meier and O'Toole (2013a) provide evidence that superintendents' answers do not reflect sophisticated adjustments based on context.

3 We chose to examine these six variables on the basis of the strong bias they appeared to produce in Meier and O'Toole's (2013b) study. Specifically, we began by identifying the seven variables in the leadership/practices section of Meier and O'Toole's (2013b) Appendix 2 that they reported produced a difference in *t*-scores with an absolute value greater than 2 (between the perceptual and archival TAKS equations). Of these seven variables, one item produced a significant effect on the TEA-reported TAKS pass rate, so we chose not to analyze this item ("With the people I have in this district, we can make virtually any program work").

school performance paired with one perceptual independent variable as well as control variables for % black students, % Latino students, % low income students, average teacher salary, student-teacher ratio, average teacher experience, teacher turnover, and a year fixed effect. Table 1 shows the resulting *t*-scores for the perceptual independent variables. Each of the 30 cells corresponds to a separate regression equation.

There is strong evidence in table 1 that the perceptual measures are subject to serious common source bias. The first and fourth columns show *t*-scores produced by regressions where the dependent variable is measured independently from the perceptual independent variable. Not a single *t*-score is significant at the .05 level. The second, third, and fifth columns show results when three different perceptual measures of performance are used as dependent variables. Every single *t*-score is statistically significant, indicating that use of a common source with these variables consistently produces false positives.

Despite the clear evidence of common source bias in table 1, Harman's single-factor test does not provide any indication of common source bias. Placing all nine perceptual variables (three dependent variables plus six independent variables) from table 1 into a principal-component factor analysis indicates that the first factor explains only about 31% of the variance. If one follows the customary (although arbitrary) practice of keeping any factor with an eigenvalue greater than 1, three factors are retained.[4] Given that the first factor explains a relatively small proportion of variance in the variables and that more than one factor was retained in the factor analysis, the results of Harman's single-factor test would appear to indicate that common source bias is not a serious problem with this data. This result highlights the unreliability of Harman's single-factor test.

Table 1
t-scores for Perceptual Independent Variables Paired with Various Dependent Variables

		Dependent Variable				
		Actual TAKS	Perceived TAKS	Perceived Quality of Ed.	SAT/ACT College Readiness	Perceived College Readiness
Perceptual independent variable	Control outside factors	0.50	2.26	2.22	0.58	2.49
	Buffering events	1.34	2.70	2.03	0.83	2.10
	Exploiting	0.04	2.58	5.89	1.17	4.84
	Prospecting strategy	1.65	6.01	7.88	-0.27	6.08
	Conflict resolution	1.51	4.49	6.27	0.72	4.11
	Shared culture	1.92	7.38	8.87	0.30	5.12
	Mean	1.22	4.15	5.53	0.63	4.16
						$N = 1{,}072$

4 Another method for choosing the number of factors to retain requires examining a scree plot and looking for the point at which the line appears to bend. This method appears to favor four factors. The eigenvalues for the first six factors are as follows: 2.80 for factor 1, 1.62 for factor 2, 1.24 for factor 3, 0.81 for factor 4, and 0.72 for factor 5, and 0.60 for factor 6. If separate factor analyses are run for each dependent variable (so that there are three factor analyses with seven variables each), exactly two factors produce eigenvalues greater than one (in all three analyses), and between 30% and 32% of the variance is explained by the first factor.

Brewer's Split Sample Method

In addition to running Harman's single-factor test, Jung (2013) implements Brewer's method of testing for common source bias. Brewer (2006) presents an innovative and intuitively appealing means of attempting to assess whether or not a common source is biasing results. His approach requires access to individual-level data which can then be aggregated up to the organizational level, so his method cannot be used with data sets that contain only one respondent per organization or that contain respondents from too few organizations to conduct a meaningful organizational-level analysis. Brewer's method requires producing an initial model which uses organizational-level means to measure the dependent variable. This initial model is then compared to a second model, which breaks the "linkage" between the dependent variable and the independent variable(s) by randomly splitting the sample of respondents into two separate groups and then using the organizational-level means of the respondents in one group to measure the dependent variable for the other group of respondents. In other words, independent variables measured by one set of respondents are used as predictors for a dependent variable measured with the other set of respondents. Brewer and Jung differ slightly in their implementation of the method, with Brewer conducting analysis at the individual level (although the dependent variable is measured at the organizational level) and Jung conducting analysis at the organizational level.

According to the test, the first model (which has the "linkage intact") may be vulnerable to common source bias because the same survey responses are used to measure independent and dependent variables. The second model uses survey responses from different individuals within each organization to measure the independent and dependent variables, so any common source bias should be eliminated (assuming that individuals' biases are not correlated with those of other respondents within their organization).[5] Brewer (2006) claims that "if common source and related bias is present, the first model should outperform the second." If both models produce similar R-squared values, this is taken as evidence that common source bias is not present. Additionally, Brewer and Jung both point out the relative similarity of coefficient estimates across the test's two models.

Despite the intuitive appeal of Brewer's split sample approach, it only addresses biases which function at a purely individual level. If perceptual biases are partially shaped by environments, as we suggest in equation (3), then biases may show up at an organizational level. Intuitively, there may be organizational factors which cause distinct survey respondents within the same organization to exhibit similar biases. A number of mechanisms could cause individual biases to be correlated with those of other individuals located within the same organization, including selection effects (organizations tend to hire individuals with similar views or personalities) and socialization processes (managers or organizational cultures help shape the ways employees perceive their organizations).

To test whether or not Brewer's split sample method is effective at detecting common source bias, we use a unique data set which contains both perceptual and

5 If the mean perceptual bias is the same for all organizations, the process of mean aggregation itself should eliminate most of the variation in bias, as long as the survey samples within organizations are sufficiently large.

archival measures of an organizational outcome. The superintendent surveys analyzed by Meier and O'Toole (2013b) contain only one respondent per organization, so Brewer's (2006) technique cannot be used with their data set. Instead, we analyze results from an employee survey of teachers in New York City schools in 2008 (New York City Department of Education 2014).[6] Survey results were merged with a data set containing publicly available city and state administrative school records for the 2007/8 school year, as described by Favero and Meier (2013). We use school violence as our dependent variable because both a perceptual survey measure and an independent archival measure of this variable are available (see Appendix 1).[7] We wish to test whether Brewer's technique detects bias, so we select as our main independent variable a survey item (concerning communication with parents and students) that appears to be related to the perceptual measure of violence but not the archival measure. Organizational-level mean estimates will be very imprecise when only a small number of responses are available, so we examine only schools in which there are at least 40 valid responses to the two survey items used in our models (meaning that there will be at least 20 responses for each mean estimate when the sample is split). Control variables are used to measure the racial/ethnic and gender composition of the student body as well as the percentage of students who are enrolled in special education and who receive free school lunches. We also include dummy variables to indicate whether or not each school offers instruction at the elementary, middle, and high school levels.[8] To split the sample, we randomly assigned half of the respondents in each school to the two groups.

Table 2 shows the results of our organizational-level analysis. The first column uses the archival measure of the dependent variable. It indicates that there is no apparent relationship between communication and violence (t-score = −0.07) when an independent source is used to measure violence. The second and third columns show the results of using Brewer's approach with the perceptual measure of the dependent variable. Model 1 (shown in the second column) indicates a significant, negative relationship between communication and violence. We conclude that this is probably a spurious relationship caused by common source bias given that the absolute value of the t-score for communication jumps by 2.56 between the first and second columns. Model 2 (shown in the third column) severs the direct linkage between the independent and dependent variables, but communication retains a significant, negative relationship with the perceptual measure of violence. This result supports the notion that teachers within the same school tend to exhibit similar perceptual biases. Models 1 and 2 explain virtually the same amount of variance, meaning that there is no indication of common source bias according to Brewer's method. This causes us to doubt

6 Access to individual-level data was obtained under a confidentiality agreement with the New York City Department of Education.

7 The archival measure is a standardized logarithmic transformation of the state's School Violence Index, which indicates the number of violent incidents per student in each school (New York State Education Department 2011). Incidents are weighted according to their severity. The perceptual and archival measures of violence are correlated at 0.55 when the perceptual measures are aggregated to the organizational level by taking the mean teacher response in each school.

8 These three categories are not mutually exclusive.

Table 2
Results from Brewer's Method

		Brewer's Method	
	Archival Violence Index	Model 1: Linkage Intact	Model 2: Linkage Severed
Communication	−.009	−.164**	−.115*
(perceptual measure)	(−0.07)	(−2.63)	(−2.09)
American/Alaskan	.030	.056	.075
native, %	(0.31)	(1.07)	(1.41)
Black, %	.007**	.010***	.010***
	(3.10)	(8.51)	(8.36)
Latino, %	.002	.006***	.007***
	(0.69)	(4.85)	(5.07)
Asian, %	−.002	.001	.001
	(−0.57)	(0.77)	(0.49)
Female, %	−.009	−.004	−.004
	(−1.78)	(−1.55)	(−1.61)
Special education, %	.026**	.012**	.008
	(3.19)	(2.71)	(1.84)
Free lunch	.005*	.000	.001
	(2.33)	(.33)	(0.61)
Elementary school	−.533***	−.146*	−.149*
	(−4.24)	(−2.21)	(−2.21)
Middle school	.375***	.128*	.164**
	(3.38)	(2.19)	(2.73)
High school	−.299*	.197*	.223**
	(−2.00)	(2.49)	(2.77)
(constant)	−.420	1.690***	1.550***
	(−0.81)	(6.15)	(5.81)
F-value	19.20***	32.42***	32.69***
R-sqr	.46	.59	.59
N	260	260	260

Note: t-scores in parentheses.
**p* < .05, ***p* < .01, ****p* < .001.

the reliability of Brewer's approach, given that the discrepancy between the first two columns strongly indicates that common source bias is present.

Because Brewer's method relies on randomly splitting the sample, one will generally obtain different results each time the test is run. In order to determine how much our results would vary over repeated draws, we conducted 100 random sample splits and computed the results for Model 2 for each split. The results are summarized in table 3. Compared to Model 1, Model 2 does tend to produce coefficients and *t*-scores for the communication variable that are slightly closer to zero, and the *R*-squared for the second model tends to be smaller. However, these differences are slight and inconsistent, as evidenced by the overlap between the Model 1 results and the middle 90% interval for Model 2. Further analysis (results not shown) seemed to indicate that a drop in *R*-squared tended to occur regardless of whether or not common source bias was present. Specifically, we tried running models identical to Models 1 and 2 except

Table 3
Summary of Results from 100 Separate Sample Splits Using Brewer's Method

		Model 2 (100 Trials)		
	Model 1	5th Percentile	Median	95th Percentile
Coefficient estimate (communication)	−.164	−.075	−.125	−.183
T-score	−2.628	−1.245	−2.095	−3.098
R-sqr	.590	.530	.559	.593

that the communication variable was omitted (thus eliminating any perceptual independent variables). A slight drop in *R*-squared for Model 2 (relative to Model 1) still tended to occur over the 100 trials. This suggests that *R*-squared may drop simply because a less precise estimate of the organizational-level mean is being used as the dependent variable in Model 2.[9] In addition to the organizational-level analyses we report in tables 2 and 3, we tried running an individual-level analysis, as Brewer (2006) did, and found fairly similar results (see Appendix 2). Despite its intuitive appeal, Brewer's method appears to perform rather poorly in practice, at least with the data set we examine.

Marker Variables

Kwon's (2012) article "Motivation, Discretion, and Corruption" finds that, in general, corruptibility is significantly related to a number of variables. Most notable are the negative relationships with extrinsic motivation and intrinsic motivation. Corruptibility is measured by asking individuals whether or not they believe receiving a certain sum of money is corruption. The use of a somewhat indirect measure may reduce social desirability bias to an extent, but the results may still be affected by common source bias, given that the independent variables are measured with responses from the same survey. Individuals who are less willing to deem a gift as corruption may also tend to exhibit an overall frustration with their workplace. Thus, the author's results may be reflecting respondents' overall feelings towards their workplace.

Kwon (2012) does acknowledge that his results are not definitive, and he briefly discusses common source bias and attempts to correct for it. Kwon assumes—as have others—that the bias can be controlled for with what has been called a "marker" variable (Lindell and Brandt 2000; Lindell and Whitney 2001). Yet Kwon himself identifies potential sources of bias (general opinions about moral values) distinct from what his marker variables directly measure (government integrity and trust). At the same time, given that the author includes many independent variables from the same survey, common source bias should be controlled for fairly well (Siemsen, Roth, and Oliveira 2010) if bias is constant across all questions, as Lindell and Whitney (2001) assume. However, several studies suggest that this assumption is often violated in practice (see Podsakoff et al. 2003; Williams, Hartman, and Cavazotte 2010), including Meier and O'Toole (2013b), who find

9 This effect may become negligible when a sufficiently large sample of respondents within each organization is available.

that bias is stronger for questions that are more vague and less factual. Podsakoff et al. (2003) point out that the effectiveness of the marker variable technique may depend on the cause of the common source bias and whether it is expected to equally affect all questions.

To better understand the effectiveness of the marker variable approach, we decided to try simultaneously including multiple perceptual independent variables in regressions with the same Texas superintendent data we used before. The ideal marker variable should be theoretically unrelated to the substantive variables and subject to the bias(es) expected to produce common method variance (Lindell and Whitney 2001; Podsakoff et al. 2003; Williams, Hartman, and Cavazotte 2010). Any of the six perceptual independent variables we identified in table 1 might be a good candidate for a marker variable because each one appeared to have little to no effect on actual performance but shared a strong measurement bias with the perceptual dependent variables. We tried placing all six variables simultaneously in the same equation. Table 4 shows the results for each of the five dependent variables examined in table 1. The first and fourth columns show the results when the independent (archival) measures of performance are used as dependent variables; none of the perceptual independent variables are significant at the .05 level. The remaining three columns show the results when a perceptual measure of performance is used. For the first perceptual independent variable (control of outside factors), the inclusion of the other five perceptual independent variables appears to adequately control for the perceptual bias that previously produced a spurious relationship with the dependent variable. Similar results are found for two of the other perceptual variables (buffering of events and conflict resolution). The other three perceptual independent variables (exploiting, prospecting strategy, and shared culture) still exhibit significant, positive relationships with at least two of the three dependent variables. Simply put, the inclusion of other perceptual variables (or "marker variables") does not appear to eliminate the bias for half of the perceptual independent variables. This suggests that perceptual bias can vary across questions and may be multidimensional (Williams, Hartman, and Cavazotte 2010). Our results provide reason to doubt that Kwon's inclusion of control variables measuring opinions about government honesty and integrity adequately addresses the concern that common source bias may be affecting his results.

Differencing

Oberfield's (2014) "Public Management in Time: A Longitudinal Examination of the Full Range of Leadership Theory" provides an interesting example of an article using a common source of data. He uses perceptual measures to show that leadership positively affects the workforce of public organizations. An obvious concern might be that employees who give favorable reports of managers' leadership might also tend to favorably describe other aspects of their workplace, causing one to doubt the validity of Oberfield's findings. What makes Oberfield's study unique is that he uses differenced data in his analysis. We are not aware of any literature discussing the issue of common source bias within the context of differenced variables.

The effect differencing will have on common source bias depends on whether the bias varies over time. Depending on the cause of bias, there is some reason to believe

Table 4
Results from Marker Variable Tests

	Actual TAKS	Perceived TAKS	Perceived Quality of Ed.	SAT/ACT College Readiness	Perceived College Readiness
Control outside factors	−.103	.010	−.001	.095	.030
	(−0.28)	(0.25)	(−0.02)	(0.19)	(0.73)
Buffering events	.412	.069	.034	.330	.044
	(1.10)	(1.68)	(0.94)	(0.65)	(1.06)
Exploiting	−.302	−.011	.094*	.602	.109*
	(−0.78)	(−0.26)	(2.52)	(1.15)	(2.55)
Prospecting strategy	.386	.141***	.154***	−.331	.147***
	(1.13)	(3.75)	(4.63)	(−0.71)	(3.87)
Conflict resolution	.267	.036	.071	.312	.049
	(0.59)	(0.72)	(1.61)	(0.50)	(0.98)
Shared culture	.568	.251***	.227***	−.077	.109*
	(1.22)	(4.92)	(5.05)	(−0.12)	(2.11)
Black, %	−.154***	−.005	−.002	−.038	−.003
	(−6.36)	(−1.76)	(−0.95)	(−1.15)	(−1.11)
Latino, %	−.087***	−.001	−.000	−.038*	.002
	(−6.56)	(−0.54)	(−0.29)	(−2.11)	(1.03)
Low income, %	−.225***	−.014***	−.013***	−.365***	−.020***
	(−12.22)	(−6.99)	(−7.47)	(−14.58)	(−9.92)
Teacher salary ($1000s)	.433***	.014	−.000	.297**	−.001
	(5.89)	(1.73)	(−0.06)	(2.98)	(−0.09)
Student-teacher ratio	−.385**	−.025	−.026*	.215	−.027*
	(−3.22)	(−1.91)	(−2.21)	(1.32)	(−2.03)
Teacher experience	.034	−.008	−.007	.114	.013
	(0.32)	(−0.70)	(−0.66)	(0.80)	(1.13)
Teacher turnover	−.296***	−.021***	−.017***	−.084	−.010**
	(−8.44)	(−5.40)	(−5.10)	(−1.77)	(−2.58)
Year 2009	−.976*	.112*	−.004	−1.295*	−.052
	(−2.09)	(2.18)	(−0.08)	(−2.04)	(−1.00)
(constant)	77.793***	3.088***	3.587***	24.202***	3.421***
	(18.55)	(6.69)	(8.83)	(4.25)	(7.36)
F-value	91.79***	25.02***	26.36***	55.24***	24.35***
R-sqr	.55	.25	.26	.42	.24
N	1,072	1,072	1,072	1,072	1,072

Note: *t*-scores in parentheses.
$*p < .05$, $**p < .01$, $***p < .001$.

that perceptual bias may be relatively constant for an individual over time. For example, Podsakoff et al. (2003, 883) state that "positive and negative affectivity are generally considered to be fairly enduring trait characteristics of the individual that may influence their responses to questionnaires." Other potential causes of bias (such as a transient mood state) are certainly more time-variant.

If an individual's perceptual error is constant over time, differencing a perceptual measure will eliminate the perceptual error. Building on equation (1), a differenced perceptual measure can be represented as follows:

$$\Delta perception(X) = X_t + p\_error_{Xt} - (X_{t-1} + p\_error_{Xt-1}) = \Delta X + \Delta p\_error_X \quad (5)$$

If the perceptual error does not change over time ($\Delta p\_error_X = 0$), the differenced perceptual measure is equal to the true differenced value of X. Thus, differenced perceptual variables should not be vulnerable to common source bias when respondents' biases are time-invariant. However, if biases do vary over time, differencing is unlikely to solve common source bias problems because the change in perceptual errors for one variable (Δp_error_X) may be correlated with the change in perceptual errors for a similarly measured variable.

Because it is not obvious to us whether it is generally reasonable to assume that biases are constant over time, we conducted empirical analyses to test whether or not differencing appeared to eliminate the common source bias we have observed in the Texas superintendent and New York City school databases. First, we tried repeating the analysis that we summarized in table 1 except that we differenced each of the variables.[10] The results are summarized in table 5. As in table 1, none of the perceptual independent variables are significantly related to the independent measures of performance at the .05 level. The differenced models appear to produce fewer spurious relationships when a common source of measurement is used, but there still appear to be some false positives. 5 out of 18 models find a significant relationship between the perceptual independent variable and the perceptual performance measure, compared with 18 out of 18 when the data were undifferenced. This may be due in part to the substantially lower proportion of variance explained by the differenced models. The average R-squared of the undifferenced models in table 1 was .32, compared to .06 for the differenced models.

Table 5
t-scores for Perceptual Independent Variables in Differenced Models

		Dependent Variable				
		Actual TAKS	Perceived TAKS	Perceived Quality of Ed.	SAT/ACT College Readiness	Perceived College Readiness
Perceptual independent variable	Control outside factors	1.34	0.74	1.82	0.07	0.87
	Buffering events	−0.41	0.27	0.89	1.55	2.69
	Exploiting	−0.74	1.40	0.93	−1.15	1.43
	Prospecting strategy	1.65	3.40	2.02	0.53	1.27
	Conflict resolution	−0.52	1.43	0.70	1.33	1.29
	Shared culture	−1.37	3.55	2.43	−0.11	0.62
	Mean	−0.01	1.80	1.47	0.37	1.36
						N = 262

10 We restricted the sample to include only districts in which the superintendent in 2011 indicated that he/she had been the superintendent of his/her district for more than two years because we wanted to analyze responses from the same respondent at multiple points in time. Differencing was accomplished by subtracting the values of the variables in 2009 from the values of the same variables in 2011. The 2009 dummy variable was dropped from the analysis because only one time period remained after differencing.

To test whether or not differencing eliminates the bias we found earlier with the New York City school data, we expanded our data set by adding data from 2007 and 2009. This produced a three year panel (2007–09), allowing us to have up to two differenced observations per school. We then ran models identical to those in the first two columns of table 2 except that we used differenced values of the variables (except for dummy variables) and added a dummy variable for 2008. Table 6 shows the results, which indicate that the perceptual independent variable is still significantly related to the perceptual (but not the archival) measure of violence. We also tried running differenced models with two-way fixed effects and standard errors clustered by organizational unit (as Oberfield did) and obtained similar results (not shown).[11]

Table 6
Results from Differenced Models of NYC School Violence

	Archival Violence Index	Perceptual Violence Measure
Communication (perceptual measure)	.022	−.160**
	(0.22)	(−3.13)
American/Alaskan native, %	.150	.018
	(1.17)	(0.27)
Black, %	−.002	.023
	(−0.06)	(1.62)
Latino, %	.005	.017
	(0.20)	(1.45)
Asian, %	−.014	−.003
	(−0.51)	(−0.21)
Female, %	−.026	.007
	(−1.25)	(0.61)
Special education, %	−.048*	−.003
	(−2.44)	(−0.29)
Free lunch, %	.002	−.001
	(0.71)	(−0.57)
Elementary school	−.056	−.005
	(−0.59)	(−0.11)
Middle school	.038	−.036
	(0.44)	(−0.81)
High school	−.117	.032
	(−1.09)	(0.59)
Year 2008	−.198**	−.039
	(−2.67)	(−1.02)
(constant)	.230*	−.025
	(2.19)	(−0.46)
F-value	2.00*	2.09*
R-sqr	.08	.08
N	307	307

Note: t-scores in parentheses.
$*p < .05$, $**p < .01$, $***p < .001$.

11 Because the dummy variables for elementary, middle, and high schools were time-invariant, they were dropped from the two-way fixed effects regressions.

In sum, our tests indicate that differencing does not guarantee that results will not be affected by common source bias, although it does seem to reduce the number of spurious findings with the variables we examined from the Texas superintendent data set. We cannot conclude that Oberfield's use of differencing adequately deals with the common-source-bias problem. It is also worth noting that differencing changes the interpretation of results and may make it more difficult to find relationships between variables that truly are related, especially if they change very slowly over time. Comparing tables 2 and 6, one can see that most of the control variables become insignificant under the differenced models, and the sign of the special education coefficient changes between the two models using the archival measure of the dependent variable. For variables that remain fairly constant but have measurement errors (random or correlated) which change substantially over time, differenced values may primarily reflect fluctuations in measurement error rather than true fluctuations in the underlying variable. At best, differencing as a means of reducing the possibility of common source bias is probably only an effective tool for applied public administration scholars to employ in limited research contexts (such as when there is good reason to believe that bias is fairly constant over time).

Finding an Independent Source of Data

Grissom's (2014) "Is Discord Detrimental? Using Institutional Variation to Identify the Impact of Public Governing Board Conflict on Outcomes" illustrates how careful research design can often be used to avoid measuring variables on both sides of a single regression equation with the same survey source. Grissom uses survey responses to measure both his key independent variable (board conflict) and most of his dependent variables (board effectiveness, board-executive relations, and teacher turnover), but he creatively draws on two distinct sets of surveys with different types of respondents. Thus, in most equations he measures one variable with results from a survey of school board members and measures the other variable with superintendent surveys. One model (his first teacher turnover model) does rely on a common source, but Grissom acknowledges the potential vulnerability of this model to common source bias and provides similar results from three other models which do not utilize a common source. Although it is possible that the bias of school board members could be correlated with the bias of superintendents from the same district, the fact that board members are not directly involved in the day-to-day operation of school districts leads us to believe that their perceptions are less likely to be affected by the environmental (organizational) factors which may have produced the common bias we observed among employees in our tests of Brewer's method. Grissom's analysis is also strengthened by the fact that he provides a set of models which use an archival measure as a dependent variable, allowing him to further consider whether the finding that board conflict reduces effectiveness is sensitive to his measurement approach. We believe that Grissom's article provides an excellent example of how scholars can design their research to mitigate the possibility of obtaining biased results when using perceptual measures.

Another Possible Approach: Structural Equation Modeling

Although we have addressed seven different ways that public administration scholars have recently responded to the common-source-bias problem, there is at least one additional approach that is worth mentioning. Structural equation modelling (SEM) offers approaches to dealing with measurement error that are not available using traditional econometric techniques. The intellectual history of SEM is distinct from the econometric tradition, but it has been found very useful among generic management scholars.[12] SEM is less common in public administration literature and thus we know of no examples of scholars within our field who have attempted to use SEM to remedy the common-source-bias problem. Nevertheless, there are some good reasons to consider exploring this approach. The most important potential strength of SEM is that it directly models measurement error and can be used to model common source bias (e.g., Williams, Hartman, and Cavazotte 2010). That is, there are models that will estimate the amount of variance that is attributable to a common method. Under certain scenarios, this may prove to be a very promising tool, but unfortunately, this method is also accompanied by several drawbacks. First, SEM models often perform poorly (and may fail to converge) if there are not multiple measures of a construct (variable). We attempted to run multiple SEM models with the Texas superintendent data, but they would not converge. Lack of convergence is a known issue with SEM (Kline 2011). Second, even when the models do converge, there is often uncertainty regarding the reliability of techniques which attempt to account for common source bias (Podsakoff et al. 2003; Podsakoff, MacKenzie, and Podsakoff 2012). If common source bias is multidimensional (as our results testing marker variables may suggest), an equation which models bias with a single dimension may not perform well. Third, as with other statistical remedies, when there is only one source that is measuring both the independent and dependent variables—as is common in the public administration and public management literature—there may be strong limitations to what an SEM model can do to improve estimation.

CONCLUSION

The common-method-bias problem suggests that how we measure our variables matters. Although perfect measures rarely (if ever) exist, researchers should be aware of the biases that may result when certain measures are selected and whether or not such biases can be remedied with statistical techniques. Using a common (or related) source to produce perceptual measures of behavioral or organizational variables can be problematic, and none of the statistical remedies that we explored appear to produce reliable fixes. Although our analysis has highlighted a potential weakness associated with using surveys to measure both independent and dependent variables, we do not mean to imply that archival data sources are without their own sets of potential problems. These problems merit their own discussion. However, we do believe that archival sources can often be utilized in combination with survey data in order to avoid the use of a single survey source for both independent and dependent variables.

12 For a full treatment of SEM, the authors recommend Rex B. Kline's (2011) *Principles and Practice of Structural Equation Modeling*.

Our essay identified six recently published articles (from this journal) and examined how they have begun to deal with the issue of common source bias. We found that these scholars employed (or failed to employ) a number of techniques for handling this problem. The first approach was to *ignore the problem.* As may be obvious at this point, this approach leaves the findings seriously suspect when the research design utilizes the same survey source to measure both the dependent variable and independent variable. Authors also attempted *adjusting the interpretation of variables.* Although there is nothing wrong with producing a study of perceptions, scholars should avoid testing a theory describing nonperceptual variables with a data set that can only provide reliable information about relationships among perceptions. The article we looked at retained many aspects of a study of nonperceptual variables even though it was presented as a perceptual study. Two creative attempted solutions to the common-source-bias problem were offered in one of the articles we considered: *Harman's single-factor test* and *Brewer's split sample method.* The author attempts to use these tests to justify the claim that common source bias is not present in his results. However, our examination of these methods using independent data sources that allow us to compare archival and perception measures indicates that the reliability of these methods is questionable at best. Harman's test does not appear particularly good at diagnosing the common-source-bias problem, and Brewer's method does not appear to successfully remedy the problem with the data set we use. Other attempts to remedy the common source bias problem are *adding marker variables to the model* or *differencing the variables*. Although each of these methods may have intuitive appeal, they also do not perform particularly well when we test them with our data. We also briefly discuss the prospects of using SEM to control for the common source bias, although in many cases it may be limited in its usefulness too.

Where does this leave us? Currently there appears to be one reliable remedy: *finding an independent data source* (Meier and O'Toole 2013b; Podsakoff et al. 2003). Scholars have highlighted the weaknesses of *statistical remedies*, particularly when the models are constructed using perceptual items from a single survey to measure nonperceptual variables on both sides of the equation. It appears as if the only reliable solution to the common-source-bias problem is to be found in *procedural remedies*, namely finding an independent source of data. Though not a general substitute for finding independent sources of data, there is also much that can be gained through careful survey design, which can reduce methods variance (Podsakoff, MacKenzie, and Podsakoff 2012). It is our hope that increased awareness of the common-source-bias problem will help lead to more carefully crafted research designs that avoid the use of a common source of perceptual data to measure both the independent and dependent variables in studies of behavior or organizations.

APPENDIX 1

DATA APPENDIX

Table A1
Survey Item Wording

Survey	Variable Name	Item Wording	Answer Choices
Texas superintendent survey	Perceived TAKS	Compared to similar districts, my assessment of our TAKS performance is.	5 = Excellent 4 = Above average 3 = Average 2 = Below average 1 = Inadequate
	Perceived quality of ed.	Compared to similar districts, my assessment of the overall quality of education in my district is.	
	Perceived college readiness	Compared to similar districts, my assessment of our college bound performance is.	
	Control outside factors	I strive to control those factors outside the school district that could have an effect on my organization.	4 = Strongly agree 3 = Tend to agree 2 = Tend to disagree 1 = Strongly disagree
	Buffering events	I always try to limit the influence of external events on my principals and teachers.	
	Exploiting	We continually search for new opportunities to provide services in our community.	
	Prospecting strategy	Our district is always among the first to adopt new ideas and practices.	
	Conflict resolution	Our district resolves conflicts by taking all interests into account.	
	Shared culture	Our district works to build a common identity and culture among district employees.	
New York City Teacher Survey	Perceived violence	Crime and violence are a problem in my school.	4 = Strongly agree 3 = Agree 2 = Disagree 1 = Strongly disagree
	Communication	How often during this school year have you: sent home information on services to help students or parents such as: tutoring, after-school programs, or workshops adults can attend to help their children in school?	5 = More than once a week 4 = Once a week 3 = About once a month 2 = About 3 or 4 times each year 1 = Rarely or never

APPENDIX 2

INDIVIDUAL-LEVEL ANALYSIS USING BREWER'S METHOD

Table B1
Results from Brewer's Method (Individual-Level)

		Brewer's Method	
	Archival Violence Index	Model 1: Linkage Intact	Model 2: Linkage Severed
Communication (perceptual measure)	−.003 (−0.79)	−.011*** (−5.32)	−.010*** (−4.71)
American/Alaskan native	.026* (1.99)	.050*** (7.01)	.050*** (6.67)
Black	.008*** (25.96)	.010*** (63.88)	.010*** (60.01)
Hispanic	.002*** (6.19)	.007*** (38.81)	.007*** (36.40)
Asian	−.002*** (−4.79)	.002*** (10.72)	.002*** (10.00)
Female	−.008*** (−12.61)	−.003*** (−9.21)	−.003*** (−8.64)
Special ed.	.023*** (21.69)	.013*** (22.38)	.013*** (20.98)
Free lunch	.005*** (16.80)	.001*** (5.45)	.001*** (5.20)
Elementary school	−.522*** (−31.36)	−.153*** (−16.98)	−.153*** (−15.94)
Middle school	.387*** (25.92)	.164*** (20.30)	.164*** (19.11)
High school	−.316*** (−16.80)	.310*** (30.47)	0.311*** (28.75)
(constant)	−0.451*** (−9.33)	1.115*** (42.54)	1.114*** (40.00)
F-value	1132.72***	1911.63***	1691.02***
R-sqr	.46	.59	.56
N	14,705	14,705	14,705

Note: *t*-scores in parentheses.
*$p < .05$, **$p < .01$, ***$p < .001$.

Table B2
Summary of Results from 100 Separate Sample Splits Using Brewer's Method (Individual-Level)

		Model 2 (100 Trials)		
	Model 1	5th Percentile	Median	95th Percentile
Coefficient estimate (communication)	−.011	−.010	−.011	−.012
t-score	−5.316	−4.435	−5.036	−5.542
R-sqr	.589	.555	.559	.564

REFERENCES

Baumgartner, Hans, and Jan-Benedict E.M. Steenkamp. 2001. Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research* 38:143–56.

Brewer, Gene A. 2006. All measures of performance are subjective: More evidence on US federal agencies. In *Public service performance: Perspectives on measurement and management*, ed. George A. Boyne, Kenneth J. Meier, Laurence J. O'Toole, Jr., and Richard W. Walker, 35–54. New York, NY: Cambridge Univ. Press.

Brown, Ben, and Wm Reed Benedict. 2002. Perceptions of the police: Past findings, methodological issues, conceptual issues and policy implications. *Policing: An International Journal of Police Strategies & Management* 25:543–80.

Campbell, Donald T., and Donald W. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56:81–105.

Cote, Joseph A., and M. Ronald Buckley. 1987. Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research* 24:315–18.

———. 1988. Measurement error and theory testing in consumer research: An illustration of the importance of construct validation. *Journal of Consumer Research* 14:579–82.

Doty, D. Harold, and William H. Glick. 1998. Common methods bias: Does common methods variance really bias results? *Organizational Research Methods* 1:374–406.

Favero, Nathan, and Kenneth J. Meier. 2013. Evaluating urban public schools: Parents, teachers, and state assessments. *Public Administration Review* 73:401–12.

Grissom, Jason A. 2014. Is discord detrimental? Using institutional variation to identify the impact of public governing board conflict on outcomes. *Journal of Public Administration Research and Theory* 24:289–315.

Im, Tobin, Wonhyuk Cho, Porumbescu, Greg, and Jungho Park. 2012. Internet, trust in government, and citizen compliance. *Journal of Public Administration Research and Theory* doi:10.1093/jopart/mus037

Jung, Chan Su. 2013. Organizational goal ambiguity and job satisfaction in the public sector. *Journal of Public Administration Research and Theory* doi:10.1093/jopart/mut020

Kline, Rex B. 2011. *Principles and practice of structural equation modeling*. New York: Guilford Press.

Kwon, Illoong. 2012. Motivation, discretion, and corruption. *Journal of Public Administration Research and Theory* doi:10.1093/jopart/mus062

Lance, Charles E., Dawson, Bryan, Birkelbach, David, and Brian J. Hoffman. 2010. Method effects, measurement error, and substantive conclusions. *Organizational Research Methods* 13:435–55.

Lee, Pei-Ching, Chen, Chin-Mei, and Chien-Heng Chou. 2013. Decreasing tax collectors' perceived social loafing through collaborative behaviors of taxpayers. *Journal of Public Administration Research and Theory* doi:10.1093/jopart/mut015

Lindell, Michael K., and Christina J. Brandt. 2000. Climate quality and climate consensus as mediators of the relationship between organizational antecedents and outcomes. *Journal of Applied Psychology* 85:331–48.

Lindell, Michael K., and David J. Whitney. 2001. Accounting for common method variance in cross-sectional designs. *Journal of Applied Psychology* 86:114–21.

Meier, Kenneth J., and Laurence J. O'Toole, Jr. 2013a. I think (I am doing well), therefore I am: Assessing the validity of administrators' self-assessments of performance. *International Public Management Journal* 16:1–27.

———. 2013b. Subjective organizational performance and measurement error: Common source bias and spurious relationships. *Journal of Public Administration Research and Theory* 23:429–56.

New York City Department of Education. 2014. NYC school survey. http://schools.nyc.gov/Accountability/tools/survey/default.htm (accessed February 3, 2014).

New York State Education Department. 2011. Violent and disruptive incidents - School safety & accountability. http://www.p12.nysed.gov/irs/vadir/vadir-accountability.html (accessed July 22, 2011).

Oberfield, Zachary W. 2014. Public management in time: A longitudinal examination of the full range of leadership theory. *Journal of Public Administration Research and Theory* 24:407–29.

Podsakoff, Philip M., MacKenzie, Scott B., Lee, Jeong-Yeon, and Nathan P. Podsakoff. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88:879–903.

Podsakoff, Philip M., MacKenzie, Scott B., and Nathan P. Podsakoff. 2012. Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology* 63:539–69.

Podsakoff, Philip M., and Dennis W. Organ. 1986. Self-reports in organizational research: Problems and prospects. *Journal of Management* 12:531–44.

Schein, Edgar H. 1992. *Organizational culture and leadership*. 2nd ed. San Francisco, CA: Jossey-Bass.

Sharma, Rajeev, Yetton, Philip, and Jeff Crawford. 2009. Estimating the effect of common method variance: The method-method pair technique with an illustration from TAM research. *MIS Quarterly* 33:473–90.

Siemsen, Enno, Roth, Aleda, and Pedro Oliveira. 2010. Common method bias in regression models with linear, quadratic, and interaction effects. *Organizational Research Methods* 13:456–76.

Williams, Larry J., Cote, Joseph A.,and M. Ronald Buckley. 1989. Lack of method variance in self-reported affect and perceptions at work: Reality or artifact? *Journal of Applied Psychology* 74:462–68.

Williams, Larry J., Hartman, Nathan, and Flavia Cavazotte. 2010. Method variance and marker variables: A review and comprehensive CFA marker technique. *Organizational Research Methods* 13:477–514.