

Boise State University

ScholarWorks

Public Policy and Administration Faculty
Publications and Presentations

Public Policy and Administration Program

4-2021

Gender and Performance in Public Organizations: A Research Synthesis and Research Agenda

Sanghee Park
Boise State University

This is an Accepted Manuscript of an Article published in *Public Management Review* on April 2021, available online at doi: [10.1080/14719037.2020.1730940](https://doi.org/10.1080/14719037.2020.1730940). The content of this document may vary from the final published version.

Gender and Performance in Public Organizations: A Research Synthesis and Research Agenda

Sanghee Park

Assistant Professor

sangheepark@boisestate.edu

Public Policy & Administration

Boise State University

Boise, ID

Published online: 11 March, 2020 in Public Management Review, <https://doi.org/10.1080/14719037.2020.1730940>

Abstract

This study examines the variations among empirical findings of gender effects on performance in public organizations; and identifies and discusses areas to be addressed in future research. The meta-analysis using 72 studies published between 1999 and 2017 presents evidence that greater representation of women in the workforce and more women in leadership roles have a positive effect on public organization performance. Study characteristics such as policy/service areas, geographical context, and time frames of the study affect the findings of gender effects, while the variance in measurement strategies and publication status do not make a difference in empirical evidence.

Keywords: gender, representation, diversity, performance, public organizations

Introduction

Public administration scholars have discussed gender in the context of social equity, workforce diversity, human resource management, and representative bureaucracy and democracy (e.g., Frederickson 2005; Meier and O'Toole 2006; Riccucci 2002). A substantial amount of evidence has accumulated regarding whether there actually is a gender difference, how to increase gender diversity, whether gender diversity increases efficiency of the public sector, and how this diversity works within different areas of services. However, the empirical evidence for gender differences and the effect of gender diversity in public organizations is mixed, sometimes conflicting, and largely focused within the context of the United States. This ambiguity may be related to methodological issues, such as potentially omitted variables, measurement strategies of the focal variables, and the study context within a specific time frame and geographical location. As gender effects in the public domain operate in a subtle way, it is difficult to verify them using gender-as-variable and hypothesis testing.

This article summarizes the empirical findings of previous research and identifies avenues for future efforts in the scholarship of gender and public administration. More specifically, this article examines the influence of study characteristics—measurement strategies, data scope, geographical contexts, and public service/policy areas—on the empirical findings about the gender-performance linkage. Meta-analysis is a useful technique to offer explanations and insights when independent studies with varying data and methods have produced different findings (Hunter and Schmidt 2014; Ringquist 2013; Stanley and Jarrell 1989). Our meta-analysis includes studies that examine the effect of having women in the leadership roles and/or more gender-diverse entities, while it excludes the studies that focus on gender congruence, moderative role of gender, and gender-based discrimination inside organizations.

The following sections review the literature on gender and performance; explain the empirical strategies including variables and selection criteria; and present the results from meta-regression analyses and address the robustness of the findings. These sections are followed by a discussion of the implications of our findings for understanding gender effects on performance and conclude with recommendations for future efforts.

Gender and Performance

Although we are greatly indebted to the past decades' hard work, there is no clear positive or negative relationship between gender and performance, mainly due to the methodological weaknesses that limit applicability and generalizability in different settings (e.g., Wise and Tschirhart 2000). This section overviews two literature streams and the variations in measurement of the focal variables in investigating the linkage between the two.

The first research stream focuses on gender differences in diverse organizational behaviors and outcomes in public organizations. These studies examine the effect of having women at the individual level on the outcomes/performance related to organizational behavior and program effectiveness. In the HR literature, gender is often coded as an indicator variable as a way to control any gender influence on organizational behavior and outcomes/performance. Leadership studies often suggest that women tend to adopt more democratic and participative leadership styles than men do (e.g., Eagly and Johannesen-Schmidt 2001) but not always (e.g. Funk 2015). Gender differences have been extensively studied in top administrative or decision-making positions, such as legislators, school principals, and supervisors, with an assumption that whether the person is female or male will make a difference both within organizations and in organizational outcomes.

A more recent and growing body of research on the relationship between gender and performance explores the effect of gender diversity/representation in the workforce within specific policy/service areas provided by diverse entities. Studies have focused on gender diversity/representation at the federal level (Choi 2009, 2013; Choi and Rainey 2010; Kim and Park 2015), the state level (Pitts, Jarry, Wilkins, and Pandey 2008; Bozeman and Feeney 2008), and the local level (Fox and Schuhmann 1999; Opstrup and Villadsen 2015; Park 2014). Studies in the context of educational institutions take up a large part of the research (Grissom, Nicholson-Crotty, Keiser 2012; Keiser, Wilkins, Meier, and Holland 2002; Smith 2015), but quite a number of studies concern police and firefighters in law enforcement agencies (Andrews and Miller 2013; Andrews, Ashworth, and Meier 2014; Meier and Nicholson-Crotty 2006), social workers and supervisors in welfare policy areas (Wilkins 2007; Wilkins and Keiser 2004), and employees in nonprofit organizations (AbouAssi and An 2017; Edwards 2012; Lee 2016; Lee and Sabharwal 2016). Studies on diversity and representative bureaucracy have measured gender as a continuous variable, such as the percentage of women, representation ratio, and Blau Index in the workforce.

Another obstacle posed in this research pertains to the different measurement strategies employed for public organization performance. The challenges of measuring organizational performance in public organizations have been noted in numerous studies (e.g., Boyne 2003; Boyne et al. 2006; Wise and Tschirhart 2000). Recent efforts address what constitutes performance in public organizations (Andersen, Boesen, and Pedersen 2016), and whether and how performance indicators are related to one another (Andersen and Hjortskov 2015; Song and Meier 2018). An important dimension of performance measurement is related to the nature of the data drawn from direct or indirect sources (archival or perceptual) and that based on stakeholders of evaluation (external and internal) (Andersen et al. 2016; Song and Meier 2018; Walker, Boyne, and Brewer 2010). Archival and external data, verified and audited by external stakeholders, provides standardized information on agency performance with a certain level of objectivity and reliability, although it may be vulnerable to manipulation (Walker et al. 2010). Perceptual and internal data attempt to assess the satisfaction of target groups but with the danger of inaccurate assessment (Andersen and Hjortskov 2015; Walker et al. 2010).

Diverse performance indicators are used to examine gender effects across the areas of interest in terms of organizational effectiveness. Archival data used for public performance measurement include arrest rates (Andrews and Miller 2013), teen pregnancy rates (Atkins and Wilkins 2013), chances of graduation (Atkins, Fertig, and Wilkins 2014), student achievement and test scores (Chukhray 2015; Dee 2007), time and money spent (Wilkins 2007; Wilkins and Keiser 2004), and agenda setting and legislative success (Bratton and Haynie 1999). Moreover, behavioral outcomes and cultural influence have received extensive attention and have been measured by perceptual data, such as job satisfaction (Choi 2009, 2013; Cohen and Vigoda 1999; Cooper et al. 2014; Grissom, Nicholson-Crotty, and Keiser 2012), turnover intention (Choi 2009; Groeneveld 2011; Grissom et al. 2012), public service motivation (DeHart-Davis, Marlowe, and Pandey 2006; Moynihan and Pandey 2007), organizational culture (Ashikali and Groeneveld 2015), and leadership and management style (Eagly and Johannesen-Schmidt 2001; Fox and Schuhmann 1999; Jacobson, Palus, and Bowling 2010).

The distinction in performance measurement between the representative bureaucracy literature and diversity management literature is also worthy of further evaluation and empirical scrutiny. Representative bureaucracy literature tends to focus on the outcomes that benefit minority groups to bear out the link between descriptive/passive and substantive/active representation, while diversity management studies engage more on the impact to overall organizational performance. Thus, the suggestion that total organizational performance can be eroded when better outcomes for a minority group come at the expense of others. The potential tradeoff between representation and equity is one of the key empirical and analytical questions in representative bureaucracy literature (Meier 2018). Empirical evidence from public education data suggests that minority administrators advocate benefits for minority groups to the extent that the benefits do not exceed the equity point to compensate for preexisting inequity (e.g., Keiser et al. 2002; Meier et al. 1999; Nicholson-Crotty et al. 2011). However, only a few studies have examined the distributional influence on the minority and society.

Empirical Models

To summarize and explain the variations among empirical results on whether and to what extent the independent variables would make a difference in organizational outcomes, we set up a meta-analysis regression model as follows:

$$b_j = \beta + \sum_{k=1}^k \alpha_k Z_{jk} + e_j \quad j=1, 2, 3, \dots L. \quad (1)$$

Here, b_j is the reported estimate of β of the j th study in literature comprised of L studies, β the ‘true’ value of the parameter of interest, Z_{jk} the meta-independent variable which measures relevant characteristics of an empirical study and explains its systematic variation from other results in the literature, α_k the meta-regression coefficient which reflects the biasing effect of particular study characteristics, and e_j the meta-regression disturbance term (Stanley and Jarrell 1989: 302; Harbord and Higgins 2008; Ringquist 2013).

Depending on purposes, several indicators have been developed to measure effect sizes as a dependent variable in meta-analysis: standardized mean differences, standardized regression coefficients, correlation coefficients, and proportions such as odds ratio. In this study, we use standardized regression coefficients of the estimate drawn from each original study as a measure of effect size. Our parameter of interest, b_j , the dependent variable in Equation (1), is the regression coefficient associated with the variable of gender or gender diversity/representation in the studies included in the meta-regression, standardized by using the equation, $\beta = b * SD(X)/SD(Y)$, which stands for how many standard deviations the response or outcome variable will change per a standard deviation increase of the focal variable. The variances of the estimates from standardized analyses are equal to one.

Using regression slopes as a dependent variable in meta-regression has raised statistical concerns about comparability among individual studies due to sampling errors and unequal variances across studies (Becker and Wu 2007; Hunter and Schmidt 2004; Jarrell and Stanley 1990; Ringquist 2013). Indeed, using the distributions of zero-order correlations (Fishers’ Z , and/or Cohen’s D and Hedges’ G for group difference) based on statistical significance and degrees of freedom is a more common strategy to estimate the effect of an intervention. However, standardized regression coefficients can be useful to summarize findings drawn from different methods and metrics across the studies in different contexts, which is also noted by Stanley and Jarrell (1989) who reviewed qualitative methods for research synthesis. This approach extends its application as it does not require raw data or summary statistics of the original studies (Becker and Wu 2007; Nieminen et al. 2013) and produces reasonably similar results from using zero-order correlations when several conditions are met (Becker and Wu 2007). Not without concerns, this strategy remains useful in offering findings that both allows systematic summary of studies within various areas/context, and serves an illustrative purpose to explain variations of gender-performance relationships.

Another issue that could confound the interpretation of the results is that this study uses a single effect size from each original study. Despite the potential bias and substantial loss of information (Ringquist 2013), the decision was made in consideration of the heterogeneity of empirical evidence from various areas, i.e., government, education, law enforcement, welfare, from which effect sizes calculated as well as the various ways “performance” is measured across the areas. For the purpose of comparing the heterogeneity of empirical evidence from various areas, it is important to satisfy the independence assumption despite the substantial loss of information. When multiple indicators are used to measure performance within a single study, our analysis uses the coefficient from the final estimation primarily interpreted and discussed in result and/or discussion sections to minimize subjectivity in choosing one effect size.

However, two effect sizes are included when they have estimated the effects of both parameters of interest, gender dummy and gender diversity/representation variables. The primary studies included two effect sizes are indicated in Table 1.

Variables Included

The meta-independent variables, Z_{jk} , reflect important characteristics of each study to explain the variations in the empirical results on the influence of gender in public organizations: measurement of the gender variable, measurement of the dependent variable, significance of the focal variable, sex of first author, data scope, publication status, geographical context, and policy/service areas/types. As the heterogeneity of data measurement has raised methodological concerns and debates, we include a variable on how the gender variable is measured by assigning the value of 1 for the non-binary continuous variable and 0 for the binary variable. In the primary studies contained within our sample, the gender variable is measured either by binary (dummy-coded) or non-binary (usually continuous) variables to assess or control the effect of gender difference and gender diversity/representation in the workforce. We also include a dummy variable to examine whether organizational outcome/performance is measured by perceptual/subjective data or archival/objective data influences the empirical findings of the study.

The variable indicating the significance of gender is included to examine whether average effect sizes vary systematically according to the statistical significance of the focal variable. Given that the effect size of the study is measured by regression coefficients, the relationship between gender and public organization performance could be different from the studies with high magnitude albeit non-significance. The gender significance is coded with an indicator variable of 0 for non-significance. As the dependent variable of this study is coded in a way to connote positive changes, we could interpret the sign and statistical significance of the variable in order to test two simple but important questions: first, does gender matter?, and second, how gender affects performance? The positive sign of the coefficient of the variable means that studies with smaller p-values are more likely to find positive gender effect on public organization performance.

This article attends to the possibility of whether researcher bias is involved in the studies of gender effects, i.e., whether the researchers' predisposition shapes the outcome. Despite well-established literature on researcher bias, little attention has been given to this issue in public administration scholarship (Bailey 1992). Researcher bias, or experimenter bias, based on the researcher's presumptions and expectations, may happen intentionally or unintentionally at any phase of research. This bias has been observed even in fields requiring rigorous procedures and evidence such as medical studies (William and Heike 1993). Given the focus on gender effects, this study attempts to test whether the researcher's gender influences empirical findings, related to the myriad choices involved in research design, data collection, methodology, or the publication process itself. The researcher's gender could be incorporated with various indicators such as the sex of the first author, a female-majority of authors, or research that included at least one female author. We use the sex of the first author in our meta-regression analysis on the ground that the first author usually takes the lead on the research and resolves any disagreement regarding research design and methods as well as substantive contents. The sex of the first author is identified primarily by their first name or websites in case of gender neutral names as well as by the personal pronoun used in their curriculum vitae.

As the true effects of gender-related variables are likely to be found in high quality data and research, we account for data/study quality in explaining variation across the empirical results by including the time scope and publication status of the primary studies. We expect that the validity and reliability of the research is contingent on the time frame of the study given the temporal nature of diversity effects (Oberfield 2014; Wise and Tschirhart 2000). As the effect of diversity changes over time, empirical studies that are cross-sectional or of a short time frame may be limited in assessing the gender-performance relationship. The primary studies' time scope is included in our model measured by the number of years used for their empirical analyses. If it is a cross-sectional study, the value is 1 for the variable. In addition to data quality, this article evaluates the influence of the publication status, whether the primary studies are published in peer-reviewed journals or in other formats.

The geographical context of the study is included in our model as an indicator variable, whether it is a US or non-US case, to examine the influence of national or cultural differences in estimated effects of gender. There is a great deal of concern that the empirical evidence leans towards the US. In particular, studies of the US context, where representative bureaucracy theory originated, take up the majority of the research (almost 70% of the primary studies included in our analysis). Further, the geographical location of a study case would be an outcome of a strategic choice of scholars who are interested in finding a significant effect of gender representation. Public bureaucracy is not

working in a vacuum but increasingly under pressures from the environment. Thus, recent work recognizes the importance of putting a global or cross-national context into public administration theory, as well as collecting empirical evidence from countries with different contextual and organizational characteristics (Groeneveld et al. 2015; Meier and Morton 2015; Meier and Funk 2017; Meier 2018; Schröter and Von Maravic 2015).

Of particular interest in this article is whether and how gender works differently across policy/service areas that each individual study examines. As organizational behavior is shaped by the policies, services and missions they administer, public organizations show different patterns in hiring and promotion of women across policy/service areas (Connell 2006; Guy 1993; Guy and Newman 2004). Out of five policy/service area, i.e., education, government, law enforcement, nonprofit, and social welfare), we use social welfare/human service agencies as a reference category. A negative sign of the coefficient means that the policy/service area reports lower performance compared to a reference category.

Sample of Studies and Selection Criteria

The primary literature included in this study was found through a computer-based search, Publish or Perish Version 5 (released 2016). The literature search was performed on January 17, 2016 to identify research, both published and unpublished, that used multivariate regression techniques to examine the effect of gender difference and/or gender diversity. We conducted an extensive search using the combination of the keywords: *gender, diversity, representative bureaucracy, the effect of gender [diversity OR representative bureaucracy] AND public sector AND/OR public organizations, performance OR outcomes, secondary data*. The same keyword search was performed for the webpages of journal publishers in the fields of public administration, public policy, and political science.ⁱ In addition, we searched for several research repositories to locate unpublished works or non-peer-reviewed research including book chapters, conference papers, dissertations, and government reports: Google Scholar, APA PsycNET, DSpace@MIT, ResearchGate, SSRN, and William & Mary Digital Archive. The database was updated in April 2018 to obtain studies appearing between January 2016 and December 31, 2017. Among the possible 428 hits, 72 research studies from 1999 and 2017 met our selection criteria. We excluded 213 studies due to the lack of a gender variable, 86 studies due to the lack of a regression analysis, 48 studies due to accessibility issues, and nine studies due to duplication. The search was finalized after adding nine studies from additional search engines. This resulted in a total of 72 studies, including 15 unpublished works, to be included in the meta-regression analysis in our study.

Criteria for including studies are (1) the study has a gender variable either as a primary independent variable or as a control variable in the model, (2) the study assesses the effect of gender with organizational outcome or performance, (3) the study reports coefficients, standard errors, and statistical significance from multiple regression. However, we excluded (1) research on performance of private organizations, (2) research using gender as mediating or moderating factors and examining the effect of gender congruence and gender-based discrimination within organizations, and (3) research statistically analyzed and tested with methods other than multiple regression, such as structural equation modeling.ⁱⁱ We also rejected (4) research without full access and research written in languages other than English. Table 1 provides a list of the original studies used in our analysis and their key differentiating characteristics.

[Table 1]

Meta-Regression Results

In order to assess whether the covariates of each parameters are significantly different from zero, we have estimated two models: (1) the default random-effect meta-regression model (REML) and (2) the method of moments (MOM) random effects model. The first random-effects meta-regression estimates the mean effect of the studies included considering the between-study heterogeneity (τ^2) by using restricted maximum likelihood (REML). The second model using the method of moments (MOM) estimation is presented as our sample size is relatively small and the estimation does not require assumptions about the distribution of the random effects (Ringquist 2013). Dealing with the potential risk of spurious findings (Type I error) in meta-regression, we perform the Monte Carlo permutation test proposed by Higgins and Thompson (2004). The p-values are adjusted for multiplicity based on random permutation in multivariable meta-regression. Table 2 presents the summary of meta-regression results.ⁱⁱⁱ

[Table 2]

The compounded hypotheses that the joint effect of the regressors equals zero for both models are rejected at $p < .05$, indicating that the log odd-ratio differs among the five categories of policy types after adjusting for the other covariates in each study. After adjusting for multiple testing, *significance of gender variable*, *time scopes*, *geographical contexts*, and policy type/areas of *government* and *nonprofit* turn out to be consistently significant (with *social welfare* as a reference category). Social welfare agencies that allocate resources to specific social groups show a consistent association of gender with higher performance when compared to general governments and nonprofit organizations. This supports the findings of previous studies (e.g., Keiser et al. 2002) that the relationship between gender and organizational outcomes/performance depends on agency/policy types/areas, and that the gender effect is more likely to be manifest in gender-salient, welfare-related organizations.

The parameter of *the significance of the focal variable* reported in the primary studies had a statistically significant and positive association with the impact of gender on performance. This suggests that studies reporting statistical significance of female leadership or a more gender-representative organization are likely to find a positive effect on performance with higher effect sizes. The estimates of the variable indicating the *geographical area* of the study signal that the studies in the US context are more likely to report positive gender effect. Furthermore, significant effects are observed for the quality of data measured by *the number of years* of the primary studies. More than half of the original studies use cross-sectional data rather than time-series or panel data over multiple years. Although cross-sectional data may contain a large number of observations, it does not guarantee higher reliability of the result due to the lack of temporal dimensions as well as the large sample bias that may result in false positives or Type I errors. The coefficient of the variable for time scope was positive and consistently significant in our model, which holds up to several robustness checks. The findings imply that gender effects on performance reinforce over time, although we cannot make any strong claims regarding dynamics within organizations.

Meanwhile, we were not able to reject the null hypotheses for the variables such as the use of different *measurement strategies of independent and dependent variables*, *sex of the first author*, and *publication form* indicating whether the original study has been published in a peer-reviewed journal or not. This implies that the gender effect on performance is relatively consistent regardless of the multitude of the empirical strategies and the publication status of primary studies. Our analysis could not confirm the distinctive effect of peer-reviewed journal publications, which are usually preferred in meta-analysis due to the quality of research.

Publication Bias and Influential Studies

This article includes unpublished “grey literature” to minimize the bias from publication selection (Ringquist 2013: 265); however, it is important to test the potential publication and other bias in meta-analysis given that null findings are less likely to be published or presented. Publication bias, or the file-drawer problem, is defined as “the association of publication probability with the statistical significance of study results.” (Sterne and Harbord 2004: 127). One way to detect the bias is to visualize scatterplots of the effects estimated from individual studies against standard error.

[Figure 1]

The diagonal lines represent the 95% confidence limits, which shows “the expected distribution of studies in the absence of heterogeneity or of selection biases” (Sterne and Harbord 2004: 131). The largest studies are placed at the top of the graph with standard error approaches zero. The funnel plot in Figure 1 shows no severe asymmetry. However, it does not guarantee the non-existence of publication bias. Given the large between-study heterogeneity, this study may have other small-study effects that lead to false-positive results. The Egger test and the Harbord’s modified test for small-study effects with different sets of three covariates reject the null hypothesis that there are no small study effects (Harbord, Harris, and Sterne 2009).

In order to check for outliers and influential studies, we present Figure 2 displaying a normal probability plot of the standardized predicted random effects, i.e., standardized empirical Bayes residuals, or standardized shrunken residuals. Scholars have proposed the use of normal probability plots showing Bayesian shrunken residuals to investigate the nature of heterogeneity. Bayesian methods estimate the posterior distributions for the study-specific effects with shrunken residuals calculated by weighting the raw residual with a shrinkage factor (SF) that includes estimated variances and the number of observations (Higgins et al. 2009). The figure suggests that the assumption of normal random effects is adequate, and there is no notable outlier because the largest shrunken residual is only slightly over two (Harbord and Higgins 2008; Palmer and Sterne 2016: 102). However, the plot detects only severe outliers because the normality assumption was used in generating the predictions (Harbord and Higgins 2008).

[Figure 2]

Additional tests were performed to check the robustness of the findings. The model reestimates the effect excluding the potentially influential studies with large sample size: the seven studies with sample size over 100,000 and the three studies over 300,000. The patterns of evidence are relatively consistent with our final model, except for the significance of the *sex of the first author* when excluding seven studies with sample size above 100,000. In addition, *the geographical context* has lost its significance in some of the reestimated models. We discuss this further in the following section. The results of the robustness checks is presented in Appendix 1B.

Discussion

Several interesting findings merit further discussion within the broad feminist theory and gender studies in public administration research. *What do we know about the variation across studies?* The purpose of our meta-regression is to investigate the degree to which factors of the original studies are related to gender effects. Some of the notable variations across studies are well worth considering. First of all, our meta-analysis did not reliably confirm the effect of the variance in measurement of the gender variable. In our samples, the gender variable measured by a dummy or a continuous variable found in a similar frequency. The findings also imply that studies at an individual level and a collective level do not report a systematic difference in the relationship between gender and public organization performance. Similarly, the measurement of dependent variables in the original studies has no robust or significant impact on the empirical results. This is consistent with the recent findings of Song and Meier (2018) about school performance assessment. More than a half of the studies included in our analysis have used archival performance measures from secondary data sources, such as graduation rates, test scores, pregnancy rates, wage gap, number of reports, and GDP per capita. About 40% percent of the studies have used perceptual data from surveys to measure attitudinal or behavioral changes of employees, such as job satisfaction and turnover intention. However, studies using surveys on external stakeholders, such as citizens, customers and other public service partners, were less frequent (e.g., Conner 2016; Yang 2007). Scholars have warned that using a perceptual performance measure may bias the result as it is difficult to differentiate assessments of organizational performance from environmental challenges (Andersen and Hjortskov 2015; Walker et al. 2010). Yet in this study the way performance or organizational outcomes are operationalized makes little difference in examining gender influence.

Does the type of public service make a difference in the relationship between gender and performance? The data included in this study reveals that the empirical research has been disproportionately distributed toward specific areas such as educational institutions and general government. Only a few studies on social welfare (e.g., Hsiesh and Guy 2009; Wilkins and Keiser 2004; Wilkins 2007), health (e.g., Ellwood and Garcia-Lacalle 2015; Moynihan and Pandey 2007), and nonprofits (e.g., AbouAssi and An 2017; Edwards 2012; Lee 2016; Lee and Sabharwal 2016) are included in our analysis. Although there are studies on environment and housing (e.g. Selden 1997), no research in these areas was included in this study. Our meta-regression finds a systematic difference in gender effects between the studies of social welfare agencies and those of government/nonprofits, while little difference is found among the studies of social welfare agencies, education institutions, and law enforcement agencies. Studies of educational institutions show a weaker relationship between gender and performance than those of social welfare agencies. Given the consistent evidence that the gender effect is systematically related to the organization/service types/areas, a greater effort needs to be exerted examining more diverse areas with different stakeholders and organizational settings.

Do gender effects vary across geographical context? We find modest evidence that there is a systematic difference between the US and non-US samples in our analysis, which implies potential variations across national contexts. However, the variable is marginally significant and fails to achieve statistical significance in robustness checks. The intuition behind this finding may simply demonstrate that the bias and discrimination against women in the public sphere transcends national and cultural boundaries, or few studies are included for non-US countries with severe gender discrimination. The assessment on the skewness of research toward the US context is supported by this study, showing that only about 30 percent of our samples are non-US studies. As we had to exclude a large volume of non-English studies and studies without full access, this assessment may have been biased. Therefore, it is important to accumulate knowledge about the gender effects in public organizations with different settings to increase the validity and explanatory power of public administration theory.

Should we address potential bias from researcher? The current study attempts to test the potential effect of researcher bias that mostly occurs unconsciously and unintentionally. It is difficult to measure and capture this subtlety, and we found little evidence in our data that gender identity of the researcher (measured by the sex of the first author)

systematically influenced the result. The robustness checks provide mixed evidence of gender bias in the research related to gender issues. There is reason for caution in the fact that the study does not fully conceptualize researcher bias in the models. Researcher bias in terms of gender may not be captured by the first author's gender if, for example, women being junior scholars and students tend to be clustered as second authors in multi-author projects. Also, gender bias toward positive findings may have less validity when all authors have a vested interest in finding positive effects. However, this finding calls attention to the researcher bias that has remained neglected and marginalized in efforts to synthesize research findings.

Does gender matter in the public sector? The major approaches in gender studies suggest that there is a gender difference and a positive effect of having women and gender-diverse organizations. The relationship between gender and performance has been examined with various data, measures and models in various settings. Despite the variations among studies, our meta-analysis corroborates the overall positive impact of gender on organizational outcomes/performance in the public sector. Yet we should note that our meta-analysis only includes studies on the direct relationship between gender and performance. Representative bureaucracy literature provides ample claims and empirical evidence supporting the nature of the gender effect that can be moderated by contextual variables (Groeneveld et al. 2015; Jacobson et al. 2010; Keiser et al. 2002; Meier and Morton 2015; Meier and Funk 2017; Meier 2018; Sowa and Selden 2003, among many). Keiser et al. (2002) made the seminal contribution to the literature by identifying the conditions of active gender representation, such as discretion, gender-salience, mission/socialization, hierarchy, stratification, critical mass, and professionalization. Most recently, Meier (2018) proposed an interactive framework to test contextual hypotheses about the impact of passive representation. Therefore, the same question, does gender matter, needs to be asked with interrogative pronouns such as when and how rather than whether it matters and what is affected.

Areas of Future Research

This study is an attempt to understand a wide range of empirical findings from previous efforts, and to recognize the impact of certain variables that account for variation, and to provide interesting points that bring to light implications for future research. Several important caveats of the analysis should be noted regarding the areas that further advance our knowledge of gender effects. First, we acknowledge the methodological criticism about effect sizes that may not be ideally comparable as well as omitted variables that may constitute performance and drive the results. Arbitrary decisions had to be made with respect to the choice of variables to some degree. A more complete model could include additional aspects such as gender role, administrative discretion, and service quality; however, these were not measured consistently in the primary studies. Relatedly, this study limits not only the variables included in the model but also the scope of the studies that specifically focus on the gender effect on organizational performance in the public sector. Although we have sought to reduce the bias by carefully applying the selection criteria on the search outcomes, problems associated with case selections from judgmental and technical reasons may be insurmountable. Lastly, this study did not differentiate the outcomes that benefit women from overall organizational performance, which bears an important question about distributive equity and outcomes to the literature of representative bureaucracy, diversity management, and public administration.

We could continue to enrich our understanding of gender-related effects through diverse methodological strategies, approaches, and sophisticated research designs using quality data that allow us to incorporate a longer time frame. As the effect of gender can be subtle, indirect, and slow-acting, we need more creative ways to answer similar questions to be explored in the future and to look beyond mere numbers. The literature could benefit from interpretive and narrative approaches as the effect of gender cannot be separable from implicit gender bias (Connell 2006; D'Agostino 2017). It is beyond the scope of our analysis, but there is insightful research on women leaders and gender diversity using in-depth interviews in a broader field of social science such as sociology, business administration, and educational engineering. Besides, the application of behavioral science to explore individuals' perception and organizational dynamics would advance both theory and practice of feminist public administration (e.g., Riccucci, Van Ryzin, and Li 2016; Riccucci and Van Ryzin 2017; Guul 2018). The symbolic aspects of representation and diversity management and the way they evolve over time is an important extension of the theory (Oberfield 2014; Riccucci and Van Ryzin 2017). Such efforts would provide a helpful frame for our expectations as we formulate hypotheses on representative bureaucracy and diversity management.

Given that the influence of gender depends on the ranks, positions, and types of jobs women are taking, more efforts in our field should be directed to identifying the potential conditional effect of gender by focusing on the organizational and environmental factors that affect the relationship. The impact of being a minority on organizational

outcome/performance may depend on their proportion within the organization and group dynamics (e.g., Kanter 1987), their willingness to take the representative role (e.g., Sowa and Selden 2003), and the national contexts (e.g., Meier and Morton 2015). Research on diversity management to elicit positive impact on organization in the public sector has been and will continue to be on the rise (Fernandez et al. 2015; Oberfield 2014; Riccucci 2002; Wise and Tschirhart 2000). In addition to the conditions under which women can make a difference in outcomes/performance, further inquiry into the effect of gender congruence (e.g., Dee 2007; Grissom et al. 2012; Guul 2018) and the intersection with other minority status such as race/ethnicity is needed (e.g., Bearfield 2009; Riccucci 2009).

Lastly, it is important to recognize the normative and empirical aspects of social equity, fairness and impartiality in gender research as an important subfield of public administration. We have little understanding of women's integration into governance and its impact on organizational justice even in this area of research (Bearfield 2009; Frederickson 2005; Guy 1993; Mastracci and Bowman 2015; Riccucci and Van Ryzin 2017). Feminist literature has noted gender divisions of labor, gender pay gap, structural barriers to women's advancement, and the emotive nature of public sector jobs (Dehart-Davis et al. 2006; Guy 1993; Guy and Newman 2004; Meier et al. 2006; Stivers 2002). However, more accurate assessments of gender disparities in the public workplace and how to achieve equitable representation of women in various areas and ranks would fill a much needed gap in the literature.

References

* denotes primary studies included in the meta-regression analysis.

- *AbouAssi, K. and S-H. An. (2017). "Gender representation and organizational size: examining opportunities for members' involvement in membership organizations." *Public Management Review* 19(10): 1437-1454.
- * Agyapong, E. (2017). "Representative Bureaucracy: Examining the Effects of Female Teachers on Girls' Education in Ghana." *International Journal of Public Administration* 1-13. doi:10.1080/01900692.2017.1388255.
- Andersen, L. B., A. Boesen, L. H. Pedersen. (2016). "Performance in Public Organizations: Clarifying the conceptual space." *Public Administration Review* 76: 852-862.
- Andersen, S. C. and M. Hjortskov (2015). "Cognitive Biases in Performance Evaluations." *Journal of Public Administration Research and Theory* 26(4): 647-662.
- *Andrews, R. and Ashworth, R. (2015). "Representation and Inclusion in Public Organizations: Evidence from the U.K. Civil Service." *Public Administration Review* 75: 279-288.
- *Andrews, R., Ashworth, R., Meier, K. J. (2014). "Representative bureaucracy and fire service performance." *International Public Management Journal* 17(1): 1-24.
- *Andrews, R., and Miller, K. J. (2013). "Representative bureaucracy, gender, and policing: The case of domestic violence arrests in England." *Public Administration* 91(4): 998-1014.
- *Ashikali, T., and Groeneveld, S. (2015). "Diversity management for all? An empirical analysis of diversity management outcomes across groups." *Personnel Review* 44(5): 757-780.
- *Atkins, D. N., and Wilkins, V. M. (2013). "Going beyond reading, writing, and arithmetic: The effects of teacher representation on teen pregnancy rates." *Journal of Public Administration Research and Theory* 23(4): 771-790.
- *Atkins, D. N., Fertig, A. R., Wilkins, V. M. (2014). "Connectedness and expectations: How minority teachers can improve educational outcomes for minority students." *Public Management Review* 16(4): 503-526.
- *Avellaneda, C. N. (2016). "Government performance and chief executives' intangibility assets: Motives, networking, and/or capacity?" *Public Management Review* 18(6): 918-947.
- Bailey, M. (1992). "Do Physicists Use Case Studies? Thoughts on Public Administration Research." *Public Administration Review* 52(1): 47-54.
- Bearfield, D. A. (2009). "Equity at the Intersection: Public Administration and the Study of Gender." *Public Administration Review* 69(3): 383-386.
- Becker, B. J. and M-J. Wu. (2007). "The Synthesis of Regression Slopes in Meta-Analysis." *Statistical Science* 22(3): 414-429.
- Bel, G., X. Fageda, M. E. Warner. (2010). "Is private production of public services cheaper than public production? A meta-regression analysis of solid waste and water services." *Journal of Policy Analysis and Management* 29(3): 553-577.
- *Bell, N. J. (2012). *Government performance, identity, and support for further devolution in Europe*. Unpublished baccalaureate thesis. The College of William and Mary, Williamsburg, VA.

- Boyne, G. A. (2003). "Sources of Public Service Improvement: A Critical Review and Research Agenda." *Journal of Public Administration Research and Theory* 13(3): 367-394.
- Boyne, G. A., K. J. Meier, L. J. O'Toole, R. M. Walker. (2006). *Public Service Performance: Perspectives on Measurement and Management*. Cambridge: Cambridge University Press.
- *Bozeman, B. and Feeney, M. K. (2008). "Public management mentoring: What affects outcomes?" *Journal of Public Administration Research and Theory* 19: 427-452.
- Bratton, K. A. and K. L. Haynie. (1999). "Agenda Setting and Legislative Success in State Legislature: The effects of gender and race." *Journal of Politics* 61(3): 658-679.
- *Choi, S. (2009). "Diversity in the US Federal Government: Diversity management and employee turnover in federal agencies." *Journal of Public Administration Research and Theory* 19: 603-630.
- *Choi, S. (2013). "Demographic Diversity of Managers and Employee Job Satisfaction: Empirical Analysis of the Federal Case." *Review of Public Personnel Administration* 33(3): 275-298.
- *Choi, S. and H. G. Rainey (2010). "Managing Diversity in U.S. Federal Agencies: Effects of Diversity and Diversity Management on Employee Perceptions of Organizational Performance." *Public Administration Review* 70(1): 109-121.
- *Chukhray, I. (2015). *School racial composition and teacher-student congruence*. Unpublished master's thesis. Rice University, Houston, TX.
- *Cohen, A., and E. Vigoda. (1999). "Politics and the workplace: An empirical examination of the relationship between political behavior and work outcomes." *Public Productivity and Management Review* 22(3): 389-406.
- *Conner, T. W. (2016). "Representation and collaboration: Exploring the role of shared identity in the collaborative process." *Public Administration Review* 76(2): 288-301.
- Connell, R. (2006). "Glass Ceilings or Gendered Institutions? Mapping the Gender Regimes of Public Sector Worksites." *Public Administration Review* 66(6): 837-749.
- *Cooper, C. A., Carpenter, D., Reiner, A., McCord, D. M. (2014). "Personality and job satisfaction: Evidence from a sample of street-level bureaucrats." *International Journal of Public Administration* 37(3): 155-162.
- *Crawford, E. R. and Fuller, E. J. (2015). "A dream attained or deferred? Examination of production and placement of Latino administrators." *Urban Education*: 52(10): 1167-1203.
- D'Agostino, M. J. (2017). "Changing the Narrative: The Difference Women Make in Public Administration." *Administration & Society* 49(1): 9-19.
- *Dee, T. S. (2007). "Teacher and the gender gaps in student achievement." *The Journal of Human Resources* XLII (3): 528-554.
- DeHart-Davis, L., Marlowe, J., Pandey, S. K. (2006). "Gender Dimensions of Public Service Motivation." *Public Administration Review* 66: 873-887.
- *Edwards, V. L. (2012). *Pathways to participation: An examination of the nature and extent of participation in civil society* (Unpublished doctoral dissertation). The University of Georgia, Athens, GA.
- *Egeberg, M., Gornitzka, A., Trondal, J. (2014). "People who run the European parliament: Staff demography and its implications." *Journal of European Integration* 36(7): 659-675.
- *Ellwood, S. and J. Garcia-Lacalle. (2015). "The Influence of Presence and Position of Women on the Boards of Directors: The Case of NHS Foundation Trusts." *Journal of Business Ethics* 130(1): 69-84.
- *Feeney, M. K. (2007). *Mentoring in the public and nonprofit sectors* (Unpublished doctoral dissertation). The University of Georgia, Athens, GA.
- *Fernandez, S., and Lee, H. (2016). "The transformation of the South African public service: Exploring the impact of racial and gender representation on organizational effectiveness." *Journal of Modern African Studies* 54(1): 91-116.
- Fernandez, S.; W. G. Resh, T. Moldogaziev, Z. W. Oberfield. (2015). "Assessing the Past and Promise of the Federal Employee Viewpoint Survey for Public Management Research: A Research Synthesis." *Public Administration Review* 75(3): 382-394.
- *Fluke, J. D., Corwin, T. W., Hollinshead, D. M., Maher, E. J. (2016). "Family preservation or child safety? Associations between child welfare workers' experience, positions, and perspectives." *Children and Youth Services Review* 69: 210-218.
- Fox, R., and R. Schuhmann. (1999). "Gender and Local Government: A Comparison of Women and Men City Managers." *Public Administration Review* 59(3): 231-242.
- Frederickson, H. G. (2005). "The State of Social Equity in American Public Administration." *National Civic Review* Winter: 31-38.

- *Funk, K. D., Silva, T., Escobar-Lemmon, M. (2017). "Leading toward equality: the effect of women mayors on gender equality in local bureaucracies." *Politics, Groups, and Identities*
<https://doi.org/10.1080/21565503.2017.1403932>
- Funk, K. D. (2015). "Gendered Governing? Women's Leadership Styles and Participatory Institutions in Brazil." *Political Research Quarterly* 68(3): 564 – 578.
- *Gazley, B., Chang, W. K., Bingham, L. B. (2010). "Board Diversity, Stakeholder Representation, and Collaborative Performance in Community Mediation Centers." *Public Administration Review* 70: 610-620.
- *Giauque, D., S. Anderfuhren-Biget, F. Varone. (2013). "HRM Practices, Intrinsic Motivators, and Organizational Performance in the Public Sector." *Public Personnel Management* 42(2): 123-150.
- *Grissom, J. A., J. Nicholson-Crotty, L. Keiser. (2012). "Does my boss's gender matter? Explaining job satisfaction and employee turnover in the public sector." *Journal of Public Administration Research and Theory* 22: 649-673.
- *Groeneveld, S. (2011). "Diversity and employee turnover in the Dutch public sector. Does diversity management make a difference?" *International Journal of Public Sector Management* 24(6): 594-612.
- Groeneveld, S., K. J. Meier, E. Schroter, R. Andrews. (2015). "Representative bureaucracy and public service performance: Where why and how does representativeness work." Presented at the Annual Conference of European Group for Public Administration, Toulouse, France.
- Guul, T. S. (2018). "The Individual-Level Effect of Gender Matching in Representative Bureaucracy." *Public Administration Review* 78(3): 398-408.
- Guy, M. E. (1993). "Three Steps Forward, Two Steps Backward: The Status of Women's Integration into Public Management." *Public Administration Review* 53(4): 285-292.
- Guy, M. E. and M. A. Newman. (2004). "Women's and Men's Job: Sex Segregation and Emotional Labor." *Public Administration Review* 64(3): 289-98.
- *Hall, T. E., and Moore, K. (2011). Poll workers and polling places. Paper prepared for the conference Bush V Gore, 10 years later: Election administration in the United States, Laguna Beach, CA.
- *Hamidullah, M. F., N. M. Riccucci, S. K. Pandey. (2013). "Women in City Hall: Gender Dimensions of Managerial Values." *American Review of Public Administration* 45(3): 247-262
- Harris, R., Bradburn, M., Deeks, J., Harbord, R., Altman, D., Sterne, J. (2008). "metan: fixed- and random-effects meta-analysis." *Stata Journal* 8(1): 3-28.
- Harbord, R. M. and J. P. T. Higgins. (2008). "Meta-regression in Stata." *The Stata Journal* 8(4): 493-519.
- Harbord, R. M., R. J. Harris, J. A. C. Sterne. (2009). "Updated Tests for Small-Study Effects in Meta-Analysis." *The Stata Journal* 9(2): 197-210.
- Higgins, J. P. T. and S. G. Thompson. (2004). "Controlling the Risk of Spurious Findings from Meta-regression." *Statistics in Medicine* 23: 1663-1682.
- Higgins, J. P. T., S. G. Thompson, D. J. Spiegelhalter. (2009). "A re-evaluation of random-effects meta-analysis." *Journal of the Royal Statistical Society* 172: 137-159.
- *Hsieh, C-W. and M. E. Guy. (2008). "Performance Outcomes: The Relationship Between Managing the "Heart" and Managing Client Satisfaction." *Review of Public Personnel Administration* 29(1): 41-57.
- Hunter, J. E. and F. L. Schmidt. (2014). *Methods of Meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.
- Jacobson, W. S., C. K. Palus, C. J. Bowling. (2010). "A Women's touch? Gendered Management and Performance in State Administration." *Journal of Public Administration Research and Theory* 20: 477-504.
- Jarrell, S. B. and T. D. Stanley. (1990). "A Meta-analysis of the Union-nonunion Wage Gap." *Industrial and Labor Relations Review* 44: 54-67.
- Kanter, R. M. (1987). "Men and Women of the Corporation Revisited." *Management Review* 76(3): 14-16.
- *Keiser, L. R., Wilkins, V. M., Meier, K. J., Holland, C. A. (2002). "Lipstick and logarithms: Gender, institutional context, and representative bureaucracy." *American Political Science Review* 96(3): 553-564.
- *Kim, C. (2003). "Representation and policy outcomes: Examining the linkage between passive and action representation." *Public Personnel Management* 32(4): 549-559.
- *Kim, S., and Park, S. (2017). "Diversity management and fairness in public organizations." *Public Organization Review* 17: 179-193.
- *Lee, Y. (2016). "Comparison between job satisfaction between nonprofit and public employees." *Nonprofit and Voluntary Sector Quarterly* 45(2): 295-313.
- *Lee, Y. and Sabharwal, M. (2016). "Education-job match, salary, and job satisfaction, across the public, for-profit, and nonprofit sectors: Survey of recent college graduates." *Public Management Review* 18(1): 40-64.

- *Lee, Y., and Won, D. (2014). "Trailblazing women in academia: Representation of women in senior faculty and the gender gap in junior faculty's salaries in higher educational institutions." *The Social Science Journal* 51: 331-340.
- *Lee, Y., and Won, D. (2016). "Applying representative bureaucracy theory to academia: Representation of women in faculty and administration and Title IX compliance in intercollegiate athletics." *Journal of Diversity in Higher Education* 9(4): 323-338.
- Mastracci, S. and L. Bowman (2015). "Public Agencies, Gendered Organizations: The future of gender studies in public management." *Public Management Review* 17(6): 857-875.
- Meier, K. J. (2018). "Theoretical Frontiers in Representative Bureaucracy: New Directions for Research." *Perspectives on Public Management and Governance* <https://doi.org/10.1093/ppmgov/gvy004>
- *Meier, K. J., and Nicholson-Crotty, J. (2006). "Gender, representative bureaucracy and law enforcement: The case of sexual assault." *Public Administration Review*: 850-860.
- *Meier, K. J., and Funk, K. D. (2017). "Women and public administration in a comparative perspective: The case of representation in Brazilian local governments." *Administration & Society* 49(1): 121-142.
- *Meier, K. J., S. H. Mastracci, K. Wilson. (2006). "Gender and Emotional Labor in Public Organizations: An Empirical Examination of the Link to Performance." *Public Administration Review* 66(6): 899-909.
- Meier, K. J., O'Toole, L. J., Goerdel, H. T. (2006). "Management Activity and Program Performance: Gender as Management Capital." *Public Administration Review* 66: 24-36.
- Meier, K. J. and T. S. M. Morton. (2015). "Representative bureaucracy in a cross-national context: Politics, identity, structure and discretion." *The Politics of Representative Bureaucracy: Power, Legitimacy, Performance*. B. G. Peters, Von Maravic, P. and Schröter, E. Cheltenham, Edward Elgar.
- Meier, K. J., R. D. Wrinkle, J. L. Polinard. (1999). "Representative Bureaucracy and Distributional Equity: Addressing the Hard Question." *The Journal of Politics* 61(4): 1025-1039.
- *Miller, A. R. and Segal, C. (2014). "Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence Escalation." University of Zurich, UBS International Center of Economics in Society, Working Paper No. 9. <http://dx.doi.org/10.2139/ssrn.2519470>
- *Moloney, K. (2007). *Representative bureaucracy: A cross-national analysis of gender (1996-2004)*. For workshop 2: Diverse Leaders Leading a Diverse Workforce of Leading the Future of The Public Sector: The Third Transatlantic Dialogue, University of Delaware, Newark, DE, USA.
- *Moon, K-K. (2017). "Fairness at the Organizational Level: Examining the Effect of Organizational Justice Climate on Collective Turnover Rates and Organizational Performance." *Public Personnel Management* 46(2):118-143.
- *Morabito, M. S., Pattavina, A., Williams, L. M. (2016). "Active representation and police response to sexual assault complaints." *Journal of Crime and Justice* 40(1): 20-33.
- *Morris, E. W., and Perry, B. L. (2016). "The punishment gap: School suspension and racial disparities in achievement." *Social Problems* 63: 68-86.
- *Moscovich, L., and Hecimovich, J. P. (2016). "Improving service delivery to women? Legislative and bureaucratic representation, political contestation, and subnational education outcomes in Argentina." Prepared for the Annual Meeting of *American Political Science Association*, Philadelphia, PA.
- *Moynihan, D. P., and Pandey, S. K. (2007). "The role of organizations in fostering public service motivation." *Public Administration Review* 67(1): 40-53.
- Nicholson-Crotty, J., J. A. Grissom, S. Nicholson-Crotty. (2011). "Bureaucratic Representation, Distributional Equity, and Democratic Values in the Administration of Public Programs." *The Journal of Politics* 73(2): 582-596.
- Nieminen, P., H. Lehtiniemi, K. Vähäkangas, A. Huusko, A. Rautio. (2013). "Standardised regression coefficient as an effect size index in summarising findings in epidemiological studies." *Epidemiology, Biostatistics and Public Health* 10(4): 1-15.
- Oberfield, Z. W. (2014). "Accounting for Time: Comparing Temporal and Atemporal Analyses of the Business Case for Diversity Management." *Public Administration Review* 74(6): 777-789.
- *Opstrup, N. and A. R. Villadsen (2015). "The Right Mix? Gender Diversity in Top Management Teams and Financial Performance." *Public Administration Review* 75(2): 291-301.
- Palmer, T. M. and J. A. C. Sterne. (2016). *Meta-Analysis in Stata: An Updated Collection from the Stata Journal*. The Stata Press.
- *Pedersen, M. J. (2016). "A 'Heart of Goal' and the Will to Succeed: Goal commitment and task performance among teachers in public schools." *Public Administration* 94: 75-88.

- *Pitts, D. W., Jerry, E. M., Wilkins, V. M., Pandey, S. K. (2006). "What do women want? Men, women, and job satisfaction in the public service." (Working paper). Retrieved from *The Social Science Research Network Electronic Paper Collection*.
- *Potipiroon, W. and M. T. Ford. (2017). "Does Public Service Motivation Always Lead to Organizational Commitment? Examining the Moderating Roles of Intrinsic Motivation and Ethical Leadership." *Public Personnel Management* 46(3): 211-238.
- Riccucci, N. M. (2002). *Managing Diversity in Public Sector Workforces*. Boulder, CO: Westview Press.
- Riccucci, N. M. (2009). "The Pursuit of Social Equity in the Federal Government: A Road Less Traveled?" *Public Administration Review* 69(3): 373-382.
- *Riccucci, N. M., Van Ryzin, G. G., Lavena, C. F. (2014). "Representative bureaucracy in policing: Does it increase perceived legitimacy?" *Journal of Public Administration Research and Theory* 24: 537-551.
- Riccucci, N. M., G. G. Van Ryzin, H. Li. (2016). "Representative Bureaucracy and the Willingness to Coproduce: An Experimental Study." *Public Administration Review* 76(1): 121-130
- Riccucci, N. M. and G. G. Van Ryzin (2017). "Representative Bureaucracy: A Lever to Enhance Social Equity, Coproduction, and Democracy." *Public Administration Review* 77(1): 21-30.
- Ringquist, E. J. (2013). *Meta-Analysis for Public Management and Policy*. Edited by Mary R. Anderson. San Francisco: Jossey-Bass.
- *Sabharwal, M. (2013). "From glass ceiling to glass cliff: Women in senior executive service." *Journal of Public Administration Research and Theory* 25: 399-426.
- *Skiba, R. J., Chung, C., Trachok, M., Baker, T. L., Sheya, A., Hughes, R. L. (2014). "Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion." *American Educational Research Journal* 51(4): 640-670.
- *Smith, A. E. (2015). "On the edge of a glass cliff: Women in leadership in public organizations." *Public Administration Quarterly*, 484-517.
- *Song, M. (2016). "Gender representation and student performance: Representative bureaucracy goes to Korea." *American Review of Public Administration* 1-18. <https://doi.org/10.1177/0042085915602537>
- *Sowa, J. E., and Selden, S. C. (2003). "Administrative Discretion and Active Representation: An expansion of the theory of representative bureaucracy." *Public Administration Review* 63(6): 700-710.
- Stanley, T. D. and S. B. Jarrell (1989). "Meta-Regression Analysis: A quantitative method of literature surveys." *Journal of Economic Surveys* 3(2): 161-170.
- *Stazyk, E. C., Davis Jr., R. S., Liang, J. (2012). "Examining the links between workforce diversity, organizational goal clarity, and job satisfaction." Prepared for the 2012 *Annual Meeting and Exhibition of the American Political Science Association*, New Orleans, LA.
- *Stearns, E., Bottia, M. C., Davalos, E., R. A., Mickelson, S. Moller, and L. Valentino. (2016). "Demographic Characteristics of High School Math and Science Teachers and Girls' Success in STEM." *Social Problems* 63(1): 87-110.
- Stivers, C. (2002). *Gender Images in Public Administration: Legitimacy and the Administrative State*. 2nd ed. Thousand Oaks, CA: Sage Publications.
- *Suhaeniti, and Ryu, S. (2013). "Gender, public management, and organizational performance: Evidence from Indonesian public schools." Prepared for presentation at the 2013 Korean Association for public administration's International Conference, Seoul, Korea.
- Walker, R.M.; Boyne, G.A. and Brewer, G.A. (2010). *Public Management and Performance*. Cambridge: Cambridge University Press.
- *Wamuthenya, W. R. (2009). "Gender differences in the determinants of formal and informal sector employment in the urban areas of Kenya across time." Presented at the 18th IAFFE Conference, Boston, MA.
- *Wilkins, V. M. (2006). "Exploring the causal story: Gender, active representation, and bureaucratic priorities." *Journal of Public Administration Research and Theory* 17: 77-94.
- *Wilkins, V. M., and Keiser, L. R. (2004). "Linking passive and active representation by gender: The case of child support agencies." *Journal of Public Administration Research and Theory* 16: 87-102.
- Wise, L. R. and M. Tschirhart. (2000). "Examining Empirical Evidence on Diversity Effects: How Useful Is Diversity Research for Public-Sector Managers?" *Public Administration Review* 60(5): 386-394.
- *Yang, K. (2007). "Making performance measurement relevant? Administrators' attitudes and structural orientations." *Public Administration Quarterly* 31(3/4): 342-383.

Table 1 List of Studies (in alphabetical order)

	Authors	Year	N	Time	US	Policy types
1	AbouAssi and An*	2017	82	1	0	government
2	Agyapong	2017	150	1	0	education
3	Andrews and Miller	2013	152	4	0	law enforcement
4	Andrews and Ashworth	2015	97	1	0	government
5	Andrews, Ashworth, and Meier	2014	138	3	0	law enforcement
6	Atkins and Wilkins	2013	506	5	1	education
7	Atkins, Fertig, Wilkins*	2014	4,253	1	1	education
8	Avellaneda	2016	135	1	0	government
9	Bell	2012	631	1	0	government
10	Bozeman and Feeney	2008	259	1	1	government
11	Choi	2009	291	1	1	government
12	Choi*	2012	176,537	1	1	government
13	Choi and Rainey*†	2010	150,000	1	1	government
14	Chukhray	2015	227	2	1	education
15	Cohen and Vigoda†	1999	200	1	0	welfare
16	Conner	2016	120	1	1	education
17	Cooper et al	2014	1,042	2	1	education
18	Crawford and Fuller	2015	49,945	20	1	education
19	Dee*	2007	4,426	1	1	education
20	Edwards	2012	219	1	1	nonprofit
21	Egeberg, Gornitzka, Trondal	2014	118	1	0	government
22	Ellwood and Garcia-Lacalle*	2015	316	3	0	welfare
23	Feeney	2007	1,220	1	1	government
24	Fernandez and Lee	2016	79	8	0	government
25	Funk, Silva, Escobar-Lemmon	2017	41,753	4	0	government
26	Gazley, Chang, Bingham	2010	160	1	0	government
27	Giauque, Anderfuhren-Biget, Varone	2013	3,131	1	0	government
28	Grissom, Nicholson-Crotty, Keiser	2012	33,900	2	1	education
29	Groeneveld	2011	23,145	1	0	government
30	Hall and Moore	2011	7,382	1	1	government
31	Halpin	2016	317	3	1	education
32	Hamidullah, Riccucci, Pandey	2013	1,263	1	1	government
33	Hsieh and Guy	2008	44	1	1	welfare
34	Keiser, Wilkins, Meier, Holland	2002	1,657	3	1	education
35	Kim*	2003	21	3	1	government
36	Kim and Park	2015	376,000	0	1	government
37	Lee	2016	339	1	1	nonprofit
38	Lee and Sabharwal	2014	3,210	1	1	nonprofit
39	Lee and Won	2014	259	1	1	education
40	Lee and Won*	2016	342	1	1	education
41	Meier and Funk	2017	5,456	1	0	government
42	Meier, Mastracci, Wilson	2006	3,118	3	1	education
43	Meier and Nicholson-Crotty	2006	449	8	1	law enforcement
44	Miller and Segal	2014	15,319	11	1	law enforcement
45	Moloney	2007	208	9	0	government
46	Moon	2017	109	4	1	government
47	Morabito, Pattavina, Williams	2016	152	2	1	law enforcement
48	Morris and Perry	2016	16,248	3	1	education
49	Moscovich and Polga-Hecimovich	2016	1,760,795	9	0	education

50	Moynihan and Pandey	2007	237	1	1	welfare
51	Nyiri and Vengroff	2005	1,000	2	0	government
52	Opstrup and Villadsen	2015	442	5	0	government
53	Park	2013	25	25	0	government
54	Park	2014	507	10	1	government
55	Pedersen	2016	7,242	1	0	education
56	Pitts, Jarry, Wilkins	2006	246	2	1	government
57	Potipiroon and Ford	2017	196	1	0	government
58	Riccucci, Van Ryzin, Lavena	2014	789	1	1	law enforcement
59	Sabharwal	2013	177,586	1	1	government
60	Skiba et al.	2014	43,320	1	1	education
61	Smith	2015	911	3	1	education
62	Song*	2016	8,655	3	0	education
63	Sowa and Selden	2003	203	1	1	government
64	Stazyk, Davis, Jiaqi	2012	433,882	3	1	government
65	Stearns et al.	2016	5,270	1	1	education
66	Suhaeniti and Ryu	2013	1,138	1	0	government
67	Wamuthenya	2009	3,238	1	0	government
68	Wang	1999	450	1	0	education
69	Wang and Yang	2015	536	1	0	government
70	Wilkins	2007	46	1	1	welfare
71	Wilkins and Keiser	2004	64	4	1	welfare
72	Yang	2007	322	1	1	government

* Two effect sizes are used from the study.

† Sample size was not reported in the result table. A reasonable sample size identified in the text is used.

Table 2 Summary of Meta-Regression Results

VARIABLES	(1) Random Effects (REML)	(2) Random Effects (GMM)
Measure of gender variable (non-binary)	-1.307 (5.033)	-1.319 (4.321)
Measure of outcome variable (Objective)	6.204 (5.143)	6.241 (4.425)
Significance of gender variable	8.493* (4.763)	8.414** (4.062)
Sex of the first author (Female)	4.604 (4.787)	4.575 (4.116)
Number of years (Time Scope)	1.498** (0.704)	1.495** (0.613)
Publication form (Peer-reviewed journal)	-4.761 (5.492)	-4.774 (4.716)
Geographical context (US)	9.255* (5.011)	9.247** (4.305)
Education	-14.71 (9.797)	-14.77* (8.409)
Government	-21.32** (9.586)	-21.27** (8.231)
Law Enforcement	-1.762 (14.09)	-1.435 (12.18)
Nonprofit	-24.16** (11.69)	-24.09** (10.04)
Constant	22.78** (10.82)	22.82** (9.291)
<i>Model Specification</i>		
Observations	81	81
Between-study heterogeneity	280.7	203.5
Adjust R-squared	20.07%	—
Prob > F	0.0134	0.0011

- 1) *** p<0.01, ** p<0.05, * p<0.1; Proportion of between-study variance explained; Joint test for all covariates with Knapp-Hartung modification.
- 2) Observations 81 from 63 studies for one effect size and nine studies with two effect sizes.
- 3) The unconditional average effect sizes from the baseline results are 23.81 for the random effects model using restricted maximum likelihood (REML) and 23.54 in the random effects model using generalized method of moments (GMM), which are both significant at the 99% confidence intervals. The predictive margins estimating the inverse-variance weighted, pooled effect sizes are 25.46 and 25.44, respectively.

Figure 1 Funnel plots: Effects and sample sizes

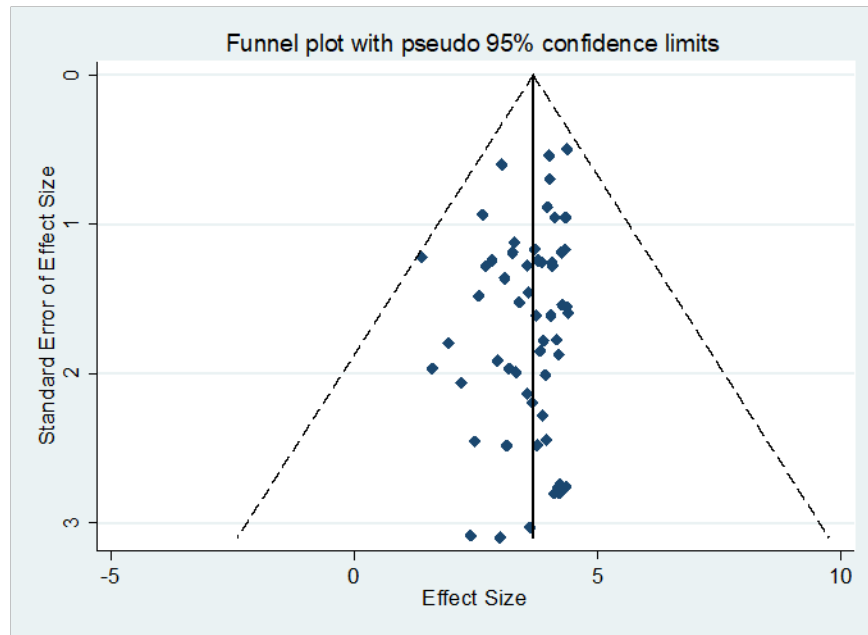
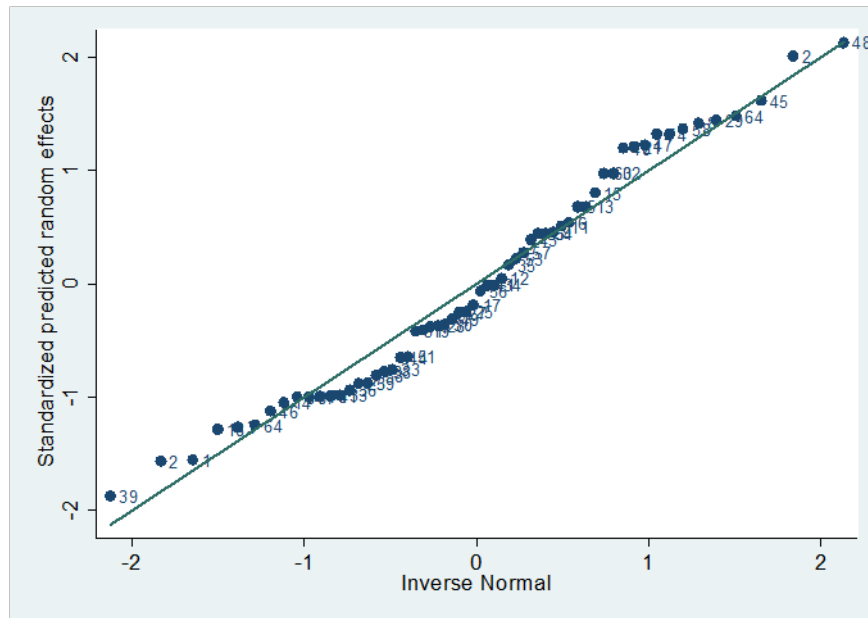


Figure 2 Normal probability plot of standardized shrunken residuals



Appendix 1A Descriptive Statistics and Correlations

	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) ES	22.65	19.27	0.03	59	1										
(2) Gen	.548	.500	0	1	.0697	1									
(3) Dep	.551	.501	0	1	.1788	.3105	1								
(4) Sig	.658	.477	0	1	.1926	.0189	-.0328	1							
(5) Sex	.500	.503	0	1	.1482	-.0514	.0957	-.0853	1						
(6) Time	2.846	3.968	0	25	.2893	.2281	.3042	.0943	.1783	1					
(7) Pub	.808	.397	0	1	-.0613	.0781	-.0487	.0584	-.1462	.0028	1				
(8) US	.615	.490	0	1	.2139	.1522	-.0499	-.0310	.0945	-.0882	.1515	1			
(9) Edu	.333	.474	0	1	.1875	.2181	.3224	.0667	-.0431	.0308	.0630	.2252	1		
(10) Gov	.423	.497	0	1	-.2123	-.1855	-.1855	.0152	-.0216	-.0078	-.2170	-.2225	-.6198	1	
(11) Law	.077	.268	0	1	.2006	.1607	.0666	.1036	-.0069	.1510	.1329	.0259	-.1969	-.2486	1
(12) NP	.064	.247	0	1	-.1596	-.1786	-.1786	-.0315	.0448	-.1158	-.0113	-.0115	-.1785	-.2254	-.0716

(1) Effect size; (2) Measurement of gender variable (0=binary, 1=non-binary); (3) Measurement of dependent variable (0=subjective, 1=objective); (4) Significance of the gender variable (0=non-significant, 1=significance); (5) Sex of the first author (0=male, 1=female) ; (6) Number of years (time-span of the study) (1=cross-sectional study); (7) Publication form (0=journal articles, 1=unpublished); (8) Geographical context (non-US=0, US=1); (9) Education; (10) Government (national/federal, state and local); (11) Law enforcement; (12) Nonprofit.

Appendix 1B Robustness Checks

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Excluding 7 studies with sample size above 100,000	Excluding 7 studies with sample size above 100,000	Excluding 3 studies with sample size above 300,000	Excluding 3 studies with sample size above 300,000	Including sample size as a covariate	Including sample size as a covariate
	Random Effects (REML)	Random Effects (GMM)	Random Effects (REML)	Random Effects (GMM)	Random Effects (REML)	Random Effects (GMM)
Measure of gender variable (non-binary)	-1.133 (5.268)	-1.166 (4.387)	-0.538 (5.018)	-0.571 (4.235)	-0.577 (4.960)	-0.589 (4.438)
Measure of outcome variable (Objective)	2.605 (5.248)	2.634 (4.391)	6.586 (5.153)	6.609 (4.364)	4.506 (5.149)	4.525 (4.615)
Significance of gender variable	8.826* (4.841)	8.685** (4.035)	3.284* (1.823)	3.236** (1.541)	7.656 (4.702)	7.607* (4.209)
Sex of the first author (Female)	10.64** (5.093)	10.63** (4.257)	7.302 (4.841)	7.258* (4.097)	5.382 (4.723)	5.364 (4.231)

Number of years (Time Scope)	1.393* (0.713)	1.398** (0.612)	1.591** (0.704)	1.607** (0.609)	1.692** (0.701)	1.695*** (0.634)
Publication form (Peer-reviewed journal)	-2.315 (5.944)	-2.323 (4.951)	-5.155 (5.767)	-5.174 (4.868)	-7.094 (5.565)	-7.095 (4.979)
Geographical context (US)	8.114 (5.207)	8.082* (4.343)	4.933 (5.100)	4.924 (4.312)	8.193 (4.960)	8.186* (4.440)
Education	-12.61 (9.556)	-12.69 (7.953)	-14.34 (9.643)	-14.40* (8.135)	-12.88 (9.680)	-12.92 (8.660)
Government	-18.91* (9.543)	-18.83** (7.951)	-24.14** (9.408)	-24.07*** (7.943)	-21.14** (9.413)	-21.11** (8.423)
Law Enforcement	-1.313 (13.71)	-0.895 (11.56)	-1.116 (13.45)	-0.759 (11.48)	-2.082 (13.85)	-1.896 (12.45)
Nonprofit	-26.29** (11.36)	-26.19*** (9.458)	-24.17** (11.30)	-24.08** (9.537)	-24.47** (11.48)	-24.42** (10.27)
Constant	20.60* (10.96)	20.65** (9.137)	25.59** (10.75)	25.62*** (9.081)	25.65** (10.75)	25.66*** (9.624)
Sample Size	—	—	—	—	-1.66e-05* (9.80e-06)	-1.66e-05* (8.77e-06)
<i>Model Specification</i>						
Observations	75	75	79	79	82	82
Between-study heterogeneity	263.5	203.5	265.8	166.7	270.6	216.3
Adjust R-squared	26.03%	—	26.39%	—	22.96%	—
Prob > F	0.0059	0.0011	0.0035	0.0002	0.0081	0.0013

*** p<0.01, ** p<0.05, * p<0.1; Proportion of between-study variance explained; Joint test for all covariates with Knapp-Hartung modification.

Notes

ⁱ Administration & Society, American Review of Public Administration, International Journal of Public Administration, International Public Management Journal, Journal of Policy Analysis and Management, Journal of Public Administration Research and Theory, Public Administration, Public Administration Review, Public Management Review, Public Personnel Management, Public Productivity and Management Review, Public Works Management & Policy, Review of Public Personnel Administration, State and Local Government Review, and others in academic publishers such as Cambridge University Press, EBSCO host, Emerald, ISI Web of Knowledge, JSTOR, Oxford Academic, ProQuest, Sage, Springer, Taylor & Francis, Wiley journal list, and WorldCat.

ⁱⁱ As our focus is on the relationship between the dependent variable and independent variables, we chose to limit the empirical modeling of the primary studies to multiple regression which is based on the linear combinations of independent variables. Unlike multiple regression estimating in sequential steps, structural equation modeling (SEM) evaluates all the variables in the model simultaneously and measurement errors are not aggregated in an error term.

ⁱⁱⁱ The baseline results are not provided as a tabular form but summarized in note 3 in Table 2. Unlike meta-regression studies examining intervention effects (e.g., an administration of new medications or an adoption of performance systems, Gerrish 2015), the unconditional average effect is of less interest of this study examining whether and how the covariates influence gender effect.