Organizational Size and Public Service Performance: An Integration of Research Findings

Richard M. Walker
Laboratory for Public Management and Policy
Department of Public Policy
City University of Hong Kong
Hong Kong
rmwalker@cityu.edu.hk

Rhys Andrews
School of Business
Cardiff University
Wales, UK
andrewsr4@cardiff.ac.uk

Bert George
Department of Public Administration
Erasmus University Rotterdam
the Netherlands
george@fsw.eur.nl

Xuan Tu
Department of Public Administration
Jiangsu Provinical Party School
PR China
vivianpink912@163.com

Conference draft: Not for quotation

Funding: National Research Foundation of Korea Grant, the Korean Government (NRF-2011-330-B00194)
Organizational Size and Public Service Performance: An Integration of Research Findings

Abstract

The optimal organizational size required to achieve desired outcomes has long been debated in public administration. Nevertheless, theory offers conflicting advice on organizational size and empirical results are often equivocal. This article seeks to bring clarity to this debate by conducting a meta-analysis of studies scrutinizing the relationship between organization size and public service performance published in the public administration literature between 2000 and 2013. Data are taken from 63 articles (226 samples) in the public administration section of the ISI Web of Science. The meta-analytical results show a positive but weak statistical relationship between organizational size and public service performance. Further analysis of substantive, methodological and contextual moderators indicates that organizational size has a positive effect on public service performance when organizational size is a dynamic measure, datasets are longitudinal and in European countries. However, in each case, the positive size effects that we identify are substantively trivial. The implications of these weak findings for research on public service performance are discussed and a research agenda addressing nonlinear and mediated and moderated relationships between size and performance is outlined.
Introduction

The question of the optimal organizational size (OS) to achieve the best performance outcomes has long been debated in the theory and practice of public administration. These debates have been shaped by a contest between economic theories that focus on the pursuit of economies of size and scope and political theories that regard closeness and responsiveness to public service users as the raison d’être of public organizations (Sharpe 1970). Subsequent arguments from public choice theory, which emphasized the role that competitive pressure could play in enhancing efficiency and effectiveness in smaller organizations, have supplemented these perspectives (Boyne 1998). Practically, these debates have seen governments around the globe change the size of public organizations. Blom-Hansen et al. (2016) describe examples of amalgamation, reduction and elimination of local governments motivated by the issue of size in 30 countries, from the 1950s to the present day.

Theory and practice present alternative remedies to resolve the problem of the optimum size for a public organization. Yet, empirical evidence remains stubbornly ambivalent about the performance effects of OS (Blom-Hansen et al. 2016; Boyne 2003; Walker and Andrews 2015). Given the on-going theoretical and practical importance of this debate, and the lack of clear conclusions, this paper asks: does organizational size matter for public service performance (PSP)? We answer this question and contribute towards knowledge on this subject through an integration of prior empirical studies published in the field of public administration from 2000 to 2013. A meta-analysis running random effect models is undertaken on 63 empirical studies (226 samples) identified in the Thompson ISI Web of Science.

The second contribution we make to the public administration literature is to tease out more nuanced relationships between OS and PSP. We examine four substantive, three methodological and a contextual factor that may influence the OS-PSP relationship because
subtle variations in theory, method and context may be masked in the aggregate level analysis. Substantively, we ask whether empirical results are influenced by the hypotheses posited by scholars in their articles, the type of organization (single purpose vs. multi-purpose), the measurement of OS and the dimension of performance under investigation. Methodologically, we investigate the treatment of OS as a control variable, the type of performance measure and the research design, and contextually we address the country of study.

In the next section the complex construct of PSP is defined, the alternative theoretical perspectives on the relationship between OS and PSP are advanced and our research expectations for the main and moderated effects are advanced. Methods are then presented. Meta-analysis findings show a weak statistically significant relationship between OS and PSP and meta-regression analysis indicates that three moderators influence the size-performance relationship: a dynamic measure of OS; a longitudinal research design and a European study context. However, the positive size effects we identify from this further analysis are trivial in magnitude. The overall implication of these findings for public administration research is that OS is not an important variable for the attainment of PSP. Practically these findings suggest that whatever other benefits they may bring, governmental changes in and experiments with OS do not result in performance improvement. Nevertheless, OS may still be an important explanatory variable influencing the connection between other internal and external contingencies and PSP. In conclusion, we therefore present arguments and develop a research agenda based around a better understanding of the nonlinear and mediated and moderated relationships between OS and PSP.

**THEORY**

**The Public Service Performance Construct**

PSP is a complex, and somewhat controversial construct (Andersen, Boesen and Pederson...
Controversy comes from the way in which reform programs have promoted performance as a means to achieve structural and cultural changes within public organizations, with many critiques focused on the purported anti-democratic motivation behind this movement (e.g. Radin 2006). Complexity arises from the multiple and potentially conflicting goals of public organizations, and the range of possible conceptual dimensions of performance, and how all of these are evaluated by a wide range of interested stakeholders. The different types and sources of data available to measure performance are a further source of conceptual complexity (Andersen, Boesen and Pederson 2016; Walker and Andrews 2015; Walker, Boyne, and Brewer 2010). To clarify the nature of this complex issue, Figure 1 offers a visual schema of the PSP construct. The horizontal axis shows stakeholders and the vertical axis types of data, with the cells providing the performance dimensions using the example of effectiveness, as measured by student attainment in secondary school education.

PSP is a multi-dimensional construct. Public organizations produce a range of outputs and outcomes and a number of models have been advanced for classifying those outputs and outcomes (Boyne 2002; Selden and Sowa 2004; Walker, Boyne, and Brewer 2010). The 3Es model, for example, emphasizes economy, efficiency, and effectiveness, while the IOO model, stresses relationships between inputs, outputs and outcomes. Within the IOO model, key outputs include: speed, quality, and quantity, and key outcomes focus on efficiency, effectiveness, equity, and responsiveness (Boyne 2002). Walker, Boyne, and Brewer (2010) extend this logic to include broader concerns about the governance of public organizations, arguing that several additional dimensions must be included in models of PSP, such as accountability, probity, and respect for human rights. To fully capture the PSP achievements of a program or organization, multiple dimensions of performance information should occupy
the cells in Figure 1. In practice, data availability often restricts empirical studies and reports of PSP by public agencies to a limited number of performance indicators. For example, integration reviews of OS and PSP find extensive use of efficiency and effectiveness and limited examples of equity (Boyne 2003; Walker and Andrews 2015).

An often-cited characteristic of public organizations is their complex operating environment, which has been empirically shown to influence performance achievements (Andrews et al. 2005; Boyne and Meier 2013; Pettigrew, Ferlie, and McKee 1992; Thompson 1967). In particular, an array of stakeholders internal and external to public organizations have a critical influence on the management and performance of those organizations. In addition to the managers and employees within an organization, external stakeholders can include overhead political authorities, other public, private, or non-profit organizations, service users, and citizens, who all have an important viewpoint on organizational achievements. These internal or external stakeholders can be incorporated in the PSP schema shown in Figure 1. Stakeholders can exert power over public agencies in different ways, whether by granting them resources or legitimacy or mandating certain activities or conferring upon agencies an urgency to act. The perspectives of stakeholders on PSP therefore matter (Mitchell, Agle, and Wood 1997). Different stakeholders are likely to express interest in different parts of the “public services production function” and prioritize alternative performance indicators. Empirical evidence on PSP offers some insights into differences between stakeholder groups and the indicators that they value most (see for example Van Ryzin et al. 2004).

Typically, PSP indicators are drawn from stakeholders’ perceptions or from archival sources (Selden and Sowa 2004; Walker, Boyne, and Brewer 2010). Perceptual measures are useful, as they cover a wide range of non-economic dimensions of performance, while being essential to core dimensions, such as responsiveness. However, perceptual measures are
vulnerable to common methods bias problems, which is problematic when the measure of performance and its determinants are taken from a single dataset and analyzed (Jakobsen and Jensen 2015; Podsakoff et al. 2003). Archival measures are seen as more robust as they are not subjective, but they present validity concerns; they are socially constructed, as actors in the public realm make decisions about what (and what not) to record as a performance-related output or outcome. Student test scores, which are frequently used in public management performance studies, usefully illustrate these tensions. The narrow focus on exam results does not account for the broad range of “value added” outputs and outcomes associated with effective schooling, such as developing character, integrity, and empathy for others, learning citizenship skills, and promoting social integration. Finally, widespread cheating has also been found, which undermines the more objective status of archival measures (Andrews, Boyne, and Walker 2011).

The complexity of the construct suggests an ideal analysis of PSP would comprehensively describe the achievements of an organization across a range of performance dimensions using perceptual and archival data collected from stakeholders internal and external to the organization. This would ensure appropriate performance information measured in fitting ways was available to stakeholders interested in understanding what managerial, organizational and political factors are likely to result in performance improvement. However, practical considerations often limit the range of performance information that can be included in studies. In keeping with Andersen, Bossen and Pedersen (2016) we propose that scholars should carefully delineate and justify the performance dimensions, stakeholders and data types in their research. In this article, we examine the consequences for the study of the OS-PSP relationship based upon the substantive choices made by authors about the dimensions of performance included in their analysis, and methodological decisions made about the measurement of OS and PSP.
**Organizational Size and Public Service Performance**

The connection between the size and the performance of public organizations has conventionally been approached from the perspective of economic theories of public service production. These theories contend that bigger organizations are more effective than smaller ones because of the economies of scale that they are able to generate (Lewin and Minton 1986; Stigler 1958). Such scale economies are likely to take three main forms.

Firstly, in big organizations the fixed costs of service production can be spread across more units of output. Examples of such fixed costs within public organizations include the common administrative overheads, technical knowledge and equipment that are utilised to achieve organizational aims (Boyne 1995). As these fixed elements of organizational activity are used more intensively, so efficiency rises and the scope for reinvesting the resulting cost-savings in performance-enhancing innovations expands. Secondly, large organizations are assumed to be able to deploy their bigger budget in ways that are conducive to better performance. In addition to having more financial resources to invest in service development, big organizations have the purchasing power to drive down suppliers’ prices and obtain more favourable rates on credit and loans for new investments (Boyne 1995; Kraljic 1983). Thirdly, large organizations offer a wider array of opportunities for professional development than their smaller counterparts, which can enable them to attract more talented and motivated employees (Black, Noel and Wang 1999; Davies 1969). In particular, bigger organizations may be able to hire better and more experienced managers who are drawn to the opportunity to take on prestigious and influential leadership roles with such organizations (Jung 2013).

Despite the venerable pedigree of scale economy arguments, public choice theory offers a contrasting economic viewpoint on the OS-PSP relationship. On the one hand, it suggests smaller public organizations are less susceptible to self-interested “producerist” oversupply of public goods because disaggregated public service delivery systems benefit
from stronger political control and greater inter-organizational competition (Niskanen 1971). On the other hand, public choice theorists claim that scale economies are exhausted by bureaucratic congestion in large organizations as the sheer number of activities requiring co-ordination overwhelm the central administrative function (Boyne and Meier 2013). In a similar vein, the organizational sociology literature has long pointed towards the task complexity associated with co-ordinating large organizations (Rushing 1967). In particular, because the number of possible social relationships increases as an exponential function of the organization’s size (Caplow 1957), bigger organizations experience more communication problems between and within the different levels of the organizational hierarchy (Morgernstern 1951).

Arguments about OS and complexity have their counterpart in political theory, which tells us that large public organizations may also be disconnected from the citizens that they serve for similar reasons to those advance by public choice theorists and organizational sociologists. Firstly, big monopolistic organizations have little incentive to reconfigure and tailor service provision to the needs of specific niche groups who are unable to exert influence over organizational decision-making (Dahl and Tufte 1973). Secondly, big bureaucracies are generally perceived to be more remote from and less interested in the concerns of the public as whole. Again, this makes it more difficult for services users to hold organizations accountable for the quality of the public services that they provide, and implies that small organizations are more responsive to users’ needs and demands than large ones (Lowndes and Sullivan 2008). Thirdly, due to all of the problems identified above, big public organizations are less likely to be trusted by citizens, which, in turn, limits their capacity to get user “buy-in” for the development of new and innovative ways of providing services (Pestoff 2014).

**Main effects and moderators**

The main purpose of a quantitative integration of research results is to identify whether there
are generalizable relationships between key variables. However, it is also necessary to move beyond aggregate population effect size analysis to explore sources of inconsistencies in those relationships and understand if they are contingent on certain conditions (Hunter and Schmidt 1990). This can be done by examining potential moderators of the main effects under consideration. In addition to investigating the overall effect of OS on PSP, we therefore assess the likely effects of substantive (hypothesized direction, organizational goals, size measurement, performance dimension), methodological (OS as a control variable, performance measurement and research design) and contextual (country of study) moderators on the OS-PSP relationship.

Main effects
Theory and (some) evidence in support of arguments about bureaucratic congestion and user responsiveness suggested that smaller organizations would have a positive association with PSP (e.g. Boyne 1996; Denters 2002; Niskanen 1971). Nevertheless, it remains conceivable that positive scale economies can enable most big organizations to overcome the problems associated with task complexity and public mistrust. For example, a number of studies indicate that it is only the very largest organizations that suffer the effects of bureaucratic congestion – and that even these may benefit from other positive size effects (e.g. Andrews, Boyne and Mostafa 2017; Rutherford, 2016). Likewise, although big bureaucracies may struggle to engage effectively with service users, research suggests that they have the resources to introduce democratic innovations that bridge the participation gap between small and large organizations (Entwistle, Andrews and Guarneros-Meza, forthcoming; Yang and Callaghan, 2005). Hence, because big bureaucracies have greater capacity for developing systems and structures for resolving co-ordination problems and eliciting user support, on the whole, they may outperform their smaller counterparts more often than not. There may, of course, be important moderators of resource-abundance scale-based arguments, such as the
propensity for big organizations to have more complex goals that we also test. Nevertheless, on balance, we expect to find a positive relationship between OS and PSP. Bigger public organizations tend to have more financial resources, greater social and political legitimacy and stronger strategic leadership, and so are well-placed to pro-actively shape the environment in which they operate (Pfeffer and Salancik 1978). By contrast, small organization are more vulnerable to external threats, especially the perils of forced amalgamation or disbandment.

Moderator analysis

Hypothesized direction

Theory on the effect of OS on PSP offers competing hypotheses. Amalgamation into larger organizations and disaggregation into smaller units are both suggested to have a positive effect on PSP. Conflicting hypotheses and varying empirical findings may be drowned out in the aggregate meta-analysis. To tease out these contradictory directions we therefore include as the first moderator the arguments made by authors in their hypotheses or statement of expectations about the impact of OS on PSP. We expect to find that the hypothesized direction in the original studies will hold in our moderator analysis.

Organizational type

The tasks and authorities that are allocated to public organizations vary considerably within and across different policy fields (Roness, 2007). In particular, many organizations are tasked with a single public purpose (e.g. border control, welfare payments), whereas others are multi-purpose organizations responsible for providing a wide range of public services (e.g. city and county governments). According to New Public Management ideas, single-purpose organizations will perform better than multi-purpose ones because they are smaller, have clearer lines of managerial and political accountability and are less susceptible to ‘producerist’ capture (Hood 1991). Indeed, many governments across the world have
embraced these ideas in pursuit of performance improvements and have disaggregated larger organizations into smaller single-purpose entities through processes of agencification (Pollitt and Bouckaert 2004). Nevertheless, whatever the merits of the arguments for agencification, there are still sound economic reasons for anticipating that multi-purpose organizations will out-perform their single-purpose counterparts. In particular, organizations providing many different services may be able to generate more internal economies of scope than single-purpose organizations by spreading fixed costs, such as administration, ICT, buildings, estates and equipment, across multiple production units (Grosskopf and Yaisawang 1990). Similarly, internal managerial efficiencies can be achieved where principal-agent hazards are spread across multiple sub-units rather than concentrated around a single core function with more leverage to behave opportunistically (Grant, Jasmine and Thomas 1988). In keeping with the broad thrust of our main effects arguments it is therefore expected that multi-purpose organizations will be associated with higher PSP.

Size measurement

The next substantive moderator is the operationalization of OS. OS is variously measured as the number of employees, the client population, or the budget of an organization. Attempts to capture overall size effects are most easily achieved through measuring the number of employees. However, data on the number of employees is not always readily available (Light 1988). When these data are not available researchers turn to proxy measures. One of the most frequently used proxies is the number of citizens that the organization serves. The arguments presented above on the main effects of OS on PSP are expected to hold when size is measured by employees or population. An alternative measure of organization size is budget. When budget is used to operationalization the concept of OS competing arguments have been presented (Boyn 2003). The first is that more resources produce better services, albeit that they need to be managed effectively. Second, this is countered by public choice prescriptions
that bigger budgets equal bloated and inefficient organizations because these resources evaporate in the bureaucracy. The latter argument has held much sway over recent decades, and in particular in times of austerity leading to an expectation that larger budgets are negatively associated with PSP.

OS is typically measured as a static concept. However, it also has an important dynamic aspect. The rate at which organizations increase or decrease in size may have a similar relationship with performance to whether or not they are large or small in the first place – for example, the empirical literature on downsizing points to a decline in public service quality and equity as the slack resources needed to respond to environmental change are “hollowed out” (Raudla, Savi and Randma-Liiv 2015). Dynamic size effects are reflected in changes in employees and budgets, and have often be captured using measure of change in the client population (e.g. Ladd 1992). Typically, studies drawing on dynamic size measures suggest that increases in the size of an organization’s client population generate more resources for the organization and are therefore associated with improved performance (e.g. Andrews et al., 2005; though see Ladd, 1992). Hence, it is expected that dynamic measures of OS will be positively associated with PSP.

Performance dimension

OS may influence PSP in different ways such as facilitating internal processes such as outputs and efficiency, meeting of organizational goals or effectiveness and the responsiveness of the organization to its citizens and users. Much of the economic argument reviewed above suggests that larger organizations are more efficient (e.g. Stigler 1958) and effective (e.g. Jung 2013). Furthermore, OS is often assumed to have a positive relationship with outputs more generally (e.g. quantity and quality) (Boyné, 1995). By contrast, responsiveness to service users’ needs and demands are argued to be a characteristic of smaller organizations (Downs 1967). For these reasons, we anticipate that the effects of OS may vary by different
dimensions of performance, having a positive relationship with measures of output (speed, quality, quantity) and outcomes (efficiency, effectiveness), and a negative relationship with measures of responsiveness.

Organizational size as a control variable

The first methodological moderator is the use of OS as a control variable. OS is often used in studies of PSP to control for the confounding effects that can be found in observational studies. OS as a control variable is included as a moderator in our analysis because it is treated in statistical models differently to when it is an independent variable that is explicitly anticipated to influence PSP. The public administration literature has not engaged in an extensive discussion about the role of control variables and their presentation in research articles. Evidence from management studies suggests that control variables are typically not well theorized and defended when included in articles (Bernerth and Aguinis 2016). If not theorized and discussed it is unclear what effects are anticipated in multivariate models. Given the preponderance not to theorize OS when it is included in models as a control variable its relationship with PSP will be non-significant.

Performance measurement

Previous quantitative studies of management and organization and PSP outcomes find that differences in the measurement of the performance construct influences statistical findings (Jakobsen and Jensen 2015; Meier and O’Toole 2013b). The central concern of these studies is with common source bias, when independent variables (in this case organizational characteristics) and the dependent variable (PSP) are drawn from a single survey source (Podsakoff et al. 2003). In this study, OS is typically measured with secondary data sources, thus common source bias many not affect the coefficients in the studies we examine. However, a critical element of this debate focuses on the likelihood that survey respondents will overestimate the performance of their organization (Meier and O’Toole 2013a).
Improvements to research design have reduced the publication of cross-sectional survey datasets in recent years in public administration. Nevertheless, use of surveys to measure public service performance was widespread during the time period of this study. For example, Walker and Andrews (2015) reported that around a third of studies in their meta-analysis of management and performance in local government adopted this practice. We therefore undertake moderator analysis on whether performance is measured using archival or perceptual data. Given the balance of evidence on this topic it would be expected that when perceptual measures of performance are used in studies there is a greater prospect of rejecting the null hypothesis and making a Type I error. The use of archival measures of performance reduces the chance of making Type I or Type II errors and should not unduly influence the study results.

Research Design

Public administration researchers increasingly employ observational research designs with the power to detect causal relationships (Zhu 2013). Central to the identification of causal relationships is time. In observational studies, this entails the measurement of the independent variables before the dependent variable, and is accommodated by longitudinal data that also offer the opportunity to lag the measurement of the dependent variable. However, not all observational research designs implement lagged structures. Such cross-sectional research designs are only able to identify associations between variables and cannot robustly identify directional relationships. Furthermore, cross-sectional designs are more susceptible to Type I and Type II errors. Given this we expect that cross-sectional designs are not able to tease out the relationship between OS and PSP whereas in more robust longitudinal designs it is expected that relationships are more likely to be identified.

Country

Context is an important variable to consider in any study of organization, management and
performance (Meier, Rutherford and Avellaneda 2017; O’Toole and Meier 2015; Thompson 1967). Public organizations vary in scope, size and purpose in ways that reflect national institutional governance arrangements. In particular, OS in the public sector is likely to be affected by the regulatory environment. The majority of the studies examined in this article are from the Europe or the U.S.A. The regulatory environment in general is more intense in the Europe than in the U.S.A. (Löfstedt and Vogel 2001). More specifically, public organizations in Europe tend to have less autonomy, more political oversight and be more constrained by labour market and other legal regulations than in the US (Meier, Rutherford and Avellaneda 2017). For this reason, public choice arguments about the relationship between small OS and PSP may be more likely to hold in the American context, whereby public organizations generally have more room for manoeuvre to compete with one another for service users (Boyne, 1998). By contrast, the need to respond to more demanding overhead regulation may result in larger organizations being associated with higher levels of PSP in Europe, because only big bureaucracies have the necessary capacity to manage these requirements from higher levels of government.

METHODS
This study utilizes a meta-analysis, or an “analysis of analyses”, to aggregate quantitative results from different studies into a single integrated literature review. A meta-analysis is defined by Glass (1976, 3) as “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings.” Despite recent calls for more meta-analyses in public policy and management, it remains a technique which is not commonly applied in public administration (Gerrish 2016; Perry 2012).

Data
The empirical literature on organizational size and public service performance was located in the public administration journals listed in the Thompson Reuters Web of Science Social
Sciences Citation Index (SSCI). Search terms included: perform*, outcome*, responsive*, consequenc*, efficienc*, effective*, satisfac* and equit*. This procedure resulted in the identification of 7,363 papers dealing with issues of PSP for the period 2000-2013. These studies were examined in greater depth, and the following decision rules for inclusion in the project were implemented: the unit of analysis was an organization, or part thereof (thus excluding studies of individuals and groups of individuals); studies included measures of management and organization (thus excluding studies of citizens’ assessments of performance); PSP was operationalized as the dependent variable; and articles presented statistical results in the form of correlations or multiple regressions that were suitable for the application of the meta-analytic technique used in this study (see below).¹ This procedure resulted in a sample of 236 articles, containing 5,915 independent tests. Our sample has an important advantage over a narrower one focused exclusively on OS. Because OS is often used as a control variable in studies of management, organization and PSP, standard search techniques are not able to capture articles that use OS as a control.² We, therefore, examined each of the 199 articles to identify those that included a measure of OS.³ The implementation of this final decision rule resulted in 63 studies containing 226 independent tests.⁴

The review strategy that we adopted benefits from focusing on peer-reviewed journal articles that were judged to be of suitable quality for publication by editors following a blind review process, and therefore expected to meet the basic requirements of theoretical and methodological rigor. It does, however, exclude unpublished papers on management and

¹ Independent or linear associations between management and performance are considered in this review. Non-linear, mediated and moderated relationships are examined in some empirical studies (e.g., Theobald and Nicholson-Crotty 2005), but not in sufficient numbers to undertake a meta-analysis.

² For example, Andersen, Boesen and Pedersen (2016) searched Scopus using the terms “performance” and “public” and returned 794 articles of which after read resulted in 61 articles.

³ Studies that included class size as a measure of OS were excluded from our analysis because a class is a sub-unit of an organization (see Meier and O’Toole 2013a, 2013b).

⁴ Identified studies were coded by two of the authors. They coded each study independently and then shared their coding results. Disagreements were dealt with through discussion until agreement was reached. This approach was taken because of the need to identify studies that included measures of OS as a control.
performance in local government and work sponsored by government, national and global organizations, such as the OECD, with an interest in the achievements of public organizations, along with books and book chapters. This approach may lead to bias, by overstating the relationship between management and performance if articles that contain statistically significant results are more likely to be published. Hence, although estimates from other fields suggest that the magnitude of publication bias is likely to be small (Rosenthal 1991), we test for ‘file drawer’ problems in our analysis.

The studies we examine were published between 2000 and 2013. The highest number of studies (10) were published in 2013 with no studies in 2002 and 2003. All of the studies contained one sample with an average sample size or number of observations of 4,146 (sd 14,505) ranging from 40 (Andrews et al. 2011) to 143,995 (Andersen and Mortensen 2009).

The majority of the studies (110 samples, 48.7 percent) offered no stated direction or hypothesis. 97 samples (42.9 percent) hypothesized that larger organizations would have a positive relationship with PSP while 6 (2.7 percent) stated smaller was better, and a null hypothesis was offered for 13 samples (5.8 percent). Of the studies in this sample 87 (38.5 percent) examined single purpose organizations including fire brigades, police departments and school districts, and the balance of 139 samples (61.5 percent) investigating multi-purpose organizations that deliver several types of public service. Size was measured as population in 80 samples (35.3 percent), budget was used in 37 samples (16.4 percent), number of students in 52 (23 percent), employees in 26 (11.5 percent) and population change in 31 (13.7 percent). The studies drew on five dimensions of performance: outputs (21, 9.2 percent), efficiency (9, 3.9 percent), effectiveness (147, 65 percent), responsiveness (29, 12.8 percent), and indexes of multiple performance indicators (20, 8.8 percent). OS was more likely to be included in PSP as a control variable (187 samples, 82.7 percent) rather than a main independent variable (39, 17.3 percent).
The majority of the studies used secondary data to operationalize performance: archival data were used in 169 (74.8 percent) samples with the balance of 57 (25.2 percent) samples using perceptual data. Nearly two-thirds (65 percent) of samples (147) implemented research designs that sought to tease out causality, largely by introducing some semblance of time into the regression model, leaving 79 samples (35 percent) purely cross-sectional studies. Half of the studies (113, 50 percent) were conducted in Europe with 89 samples (39.4 percent) from the U.S.A. and the balance of 24 samples from other countries (10.6 percent).

**Meta-analytical procedure**

The meta-analytical procedure of Field and Gillet (2010) was used to aggregate, integrate and synthesize the findings of the empirical studies on the relationship between OS and PSP. This procedure allowed us to (a) derive a population effect size across correlations, (b) test whether specific moderating variables were significant sources of variation among effect sizes and (c) identify potential issues with publication bias. All analyses were conducted using custom-written syntax for SPSS and R (Field and Gillett 2010). This procedure and the accompanying syntax have been used and validated by a variety of studies (e.g. Brewin et al. 2007; Richardson, Abraham and Bond 2012). Below we describe the four-step approach implemented in this article.

**Identifying effect sizes**

First, the effect size for each relationship identified in a study was calculated. Correlation coefficients are a well-known and easily interpretable effect size across the social sciences (Field and Gillett 2010; Rosenthal and DiMatteo 2001) and are particularly recommended for meta-analyses in public management and policy studies (Ringquist 2013). However, many public administration articles publish multiple regression results without correlation coefficients. It was therefore necessary to calculate correlation coefficients based on the t-

---

5 Two studies reported correlation coefficients and regression coefficients, the remained regression coefficients only.
statistic and the degrees of freedom:

\[ r = \frac{t^2}{(t^2 + df)} \]

with \( t \) = the t-statistic and \( df \) = degrees of freedom

The t-statistic can also be calculated by dividing the unstandardized regression coefficient and their standard errors. Although these metrics were typically reported in the analyzed studies, 12 articles did not provide sufficient information to calculate \( r \) using the above-mentioned process. The original authors of these 12 studies were contacted and asked to provide the necessary metrics, of which ten responded. For the two remaining studies, other formulas were used to calculate the correlation coefficients (Ringquist 2013, 105-110). To maximize the number of size effects, size effects were calculated for each sample.

**Conducting the meta-analysis and subgroup analyses**

Second, a population effect size was derived and subgroup analyses carried out to assess potential moderators. A random-effects meta-analysis model using Fisher’s r-to-Z transformation was selected for the analysis undertaken in this article (Gerrish 2016; Hedges and Vevea 1998; Homberg, McCarthy and Tabvuma 2015) A random-effects model is preferred because we are working with “real-world data … [that] are likely to have variable population parameters” (Field and Gillett 2010, 673). A fixed effect model would not be appropriate given that sample characteristics vary considerably. In order to be able to generalize study findings to a broader population (i.e. not just the studies included in a meta-analysis) the unconditional inferences presented by random-effects models is required (Hedges and Vevea 1998).

Fisher’s r-to-Z transformation is a convention in meta-analysis (Ringquist 2013). To remove the slightly positive bias that results from the Fisher-transformed r’s, the effect sizes are first transformed with \( r - \frac{[(r(1 - r^2))/2(n - 3)]}{2(n - 3)} \) before the Fisher transformation is applied (Field and Gillett 2010). Following this the overall variation in effect sizes is calculated as
well as the population effect size and its 95% confidence interval. To identify how this aggregate effect size might differ across our moderators, the sample is divided into subgroups and the same procedure is undertaken within each subgroup. This twofold approach allows us to identify both the overall population effect size as well as potential moderators to include in our meta-regression analysis. The comparatively small number of studies and effect sizes included in our study limits the number of moderators that can reasonably be included in the meta-regression analysis. Conducting ex ante subgroup analyses provides us with (a) a statistical rationale to include relevant moderators in the meta-regression analysis, and (b) a means of interpreting the role of these moderators as sources of variation in effect sizes.

Conducting the meta-regression analyses

Third, a meta-regression analysis to test for differences in effect sizes across groups was performed to identify whether the moderators from the second step were significant sources of variation among effect sizes using a random-effects model. A general linear model is constructed in which the \( z \)-transformed effect size is the dependent variable and the indicated moderators from the previous step are the independent variables. This is a mixed model which employs a generalized least squared (GLS) estimate to calculate the effect of each moderator (Field and Gillett 2010).

Identifying publication bias

Last, visual and statistical tests to detect potential issues with publication bias were conducted. Pro-positive bias in the academic community – the “file draw problem” – leads to more positive research findings being published, than null results (Franco, Malhotra and Simonovits 2014; Ringquist 2013; van Witteloostuijn 2016). Hence, publication bias implies that a meta-analysis based on articles published in peer-reviewed journals is likely to overestimate population effects because these do not include unpublished articles (Field and Gillett 2010). There is no “definitive” test which can precisely estimate the presence of
publication bias (Ringquist 2013). Nonetheless, we report three specific tests for publication bias in meta-analysis as recommended by Field and Gillet (2010): Rosenthal’s Fail-Safe N, the rank correlation test of Begg and Mazumdar (1994) and a funnel plot to indicate potential asymmetry of effect sizes.

**RESULTS**

**Population effect size and subgroup analyses**

A weak statistically significant relation for the main population effect size on the relationship between OS and PSP is shown in the first row of Table 1. The estimated effect size is .006 [CI₉₅: -.001, .012], with z-value of 1.775 (p = .076). While the coefficient is marginally statistically significant the size effect is trivial. The homogeneity test indicates that there is considerable variation in effect sizes overall validating the proposed moderator analysis ($\chi^2 = 727.475$, df = 225, p = .000).

[Insert Table 1 about here]

Subgroup analysis was conducted on the moderators, of which nine were statistically significant: larger organization hypothesized to lead to higher performance (r = .039), relation not stated (r = -.005), multiple goals (r = .027), population change (r = .148), perceptual performance measurement (r = -.011), archival performance measurement (r = .007), longitudinal research design (r = .015), USA (r = -.008) and EU (r = .016). All moderator size effects are trivial with the exception of population change, for which a small to medium overall effect size is observed (Cohen 1988).

**Meta-regression analysis**

A meta-regression using generalized least squared (GLS) estimates was conducted based on the nine statistically significant moderators in the subgroup analyses (Field and Gillett, 2010). The model is:
\[ Z_i = \beta_0 + \beta_1 \text{LARGER} + \beta_2 \text{NODIR} + \beta_3 \text{MULTI} + \beta_4 \text{POPCHAN} + \beta_5 \text{ARCP} + \beta_6 \text{LONG} + \beta_7 \text{EU} + \beta_8 \text{OTHER} + e_i \]

where larger organization is hypothesized to lead to higher performance (LARGER), no direction stated (NODIR), multiple goals (MULTI), population change (POPCHAN), objective performance measurement (ARCP), longitudinal data (LONG), Europe (EU) and other countries outside the USA/EU (OTHER), and with \( e_i \) representing the within-study error variance.

The meta-regression confirms the overall weak significant and trivial effect of OS on PSP \( (r = .006; \text{CI}_{95}: -.001, .012; t = 1.768; p = .078) \). Table 2 present the coefficients of the above-mentioned equation. Four moderators are statistically significant when compared to their specific reference categories: NODIR \( (\beta = -.020) \), POPCHAN \( (\beta = .063) \), LONG \( (\beta = .016) \) and EU \( (\beta = .016) \), however NODIR and EU were significant at the weaker 0.1 level of statistical significance.

[Insert Table 2 about here]

**Publication bias**

Publication bias tests are conducted to address pro-publication bias. While our findings are weak and thus not suggestive of strong publication bias, the three tests we conduct suggest that there is a medium threat of publication bias. Rosenthal’s Fail-Safe N equals 803, which implies that 803 additional correlations would have to be published to significantly alter the findings of our meta-analysis. The rank correlation test of Begg and Mazumdar (1994) indicates a correlation between the effect size and its standard error of .339 \( (p < .01) \), which is indicative of a significant but medium threat of publication bias. Finally, the funnel plot illustrated in Figure 2 indicates an overall symmetry of effect sizes although some outliers are apparent.

[Insert Figure 2 about here]
DISCUSSION

Implication for theory

Finding a weak statistical relationship between OS and PSP is contrary to the strong theoretical arguments made on the role of OS as an antecedent of performance and our primary research hypothesis. While theory offers contrasting perspectives on the role of large and small organizations in achieving higher levels of PSP a definitive point of view is not supported. Economic arguments for higher performance effects in bigger organizations focus on the efficiency gains that could be derived from size. Our study was not able to fully confirm or reject these arguments, partly because of the limited number of studies employing efficiency as the dependent variable. However, when OS theorists present arguments that larger organizations may be able to attract more talented staff a positive relationship with effectiveness would be anticipated because more skilled employees should be able to deliver against organizational goals.

The majority of the studies examined in this article included a measure of effectiveness, but clear-cut OS-effectiveness performance effects were not detected. Public choice theorists and political theory on government-citizen relationships have made the case for small organizations being more responsive to internal and external contingencies – intra-organizational communication and coordination is more easily achieved and outreach to citizens can be established. These “softer” less tangible public service outcomes may explain the negative OS-PSP relationship associated with perceptual performance measures. The results from the meta-analysis also show a weak negative statistically significant relationship between OS and PSP when scholars do not state a direction for the OS-PSP relationship that might offer some support to the public choice arguments. However, it is important to note that this finding relates to studies that did not explicitly theorize the OS-PSP relationship, thus this can be viewed as a chance finding.
It is possible that the weak overall findings point to the arguments for positive and negative scale economies cancelling each other out. However, our moderator analysis did not confirm this. Therefore, we argue that theory on OS should not consider the OS-PSP relationship as a linear or an independent relationship.

The first way we suggest to advance theory is that if scholars wish to consider the OS-PSP relationship as an independent relationship it should be conceptualized as nonlinear, especially in an inverted U-shape indicating that there is an optimal organization size beyond which the benefits of being big are lost. This argument would be in keeping with the ideas of some public choice theorists who argue that scale economies cannot be reaped in very small or very large organizations, due to the effects of bureaucratic congestion (Andrews, Boyne and Mostafa 2017). For example, Bohte and Meier’s (2001) study of “span of control” examines school size. They argue, and provide empirical evidence, that as schools get larger they are able to offer benefits, such as higher salaries to attract higher calibre personnel but that the performance gains associated with these benefits eventually diminish. Building explicitly on that research, Theobald and Nicolson-Crotty (2005) show that larger size gradually impedes communication, coordination and supervision, and past a certain point has a detrimental effect on effectiveness and equity. Examining corporate capacity in local government, that is the effect of OS on PSP in one part of the organization, Andrews and Boyne (2011) identify an inverted u-shaped size effect across three dimensions of performance – effectiveness, cost effectiveness and equity. These findings are based upon the argument that while all organizations require central administrative services at some point these overhead costs reflect a bureaucratic burden with harmful effects on service delivery and PSP. more generally, the findings are in keeping with a growing body of scholarship in public administration that suggests that examining linear or independent associations does not capture the complexity of the relationship between management, organization and PSP,
including innovation (O’Toole and Meier, 1999; 2015; Provan and Kenis 2008; Walker, Avellaneda and Berry 2016)

One part of this agenda would be to investigate where the tipping point from positive to negative performance effects occurs. Andrews and Boyne’s (2011) study of corporate capacity notes that performance effects begin to turn south when the budget of the corporate centre reaches two percent of total expenditure – marginally more than the average corporate budget size. Theobald and Nicholson-Crotty (2005) find that when school size reaches 902 out of a maximum school size of 1149, or close to the maximum OS in the data they examined. Although it is clearly difficult to draw lessons about the right OS in the public sector from these initial empirical results, research in education economics suggests that it is possible to use meta-analysis techniques to identify the optimal of public organizations (e.g. Colegrave and Giles 2008). Hence, more studies examining nonlinear size-performance relationships could assist public administration researchers to develop new theory and evidence on the importance of OS for PSP.

Sociological theories of OS have long considered the relationship between OS and a range of management and organizational variables (Blau, 1970; Hall, Johnson and Haas, 1967; Kimberley 1976). These theories have considered OS one variable in a more complex set of inter-relationships between the environment, management, organization and outcomes – an approach that has been especially evident in the development of contingency theory within organization studies (Donaldson 2001). The notion of fit found within contingency theory suggests that organizations need to align their strategy, structure and other key contingencies, such as OS, to generate high performance (Burns and Stalker 1961; Child 1975; Thompson 1967). OS was extensively examined in the contingency theory studies of the Institute for Local Government (Greenwood, Hinings and Ranson 1975), and recent applications of Miles and Snow’s (1978) strategic management framework in the public sector indicate that this
style of theorizing continues to have great relevance for understanding the determinants of PSP (Walker 2013). A number of the studies examined in this meta-analysis do locate OS within conceptual frameworks that examine the inter-relationship between management and organization variables. However, the consideration of joint relationships between these variables and OS is scant (indeed a large number of the studies we examined did not offer any conceptual or theoretical rationale for the inclusion of OS in the statistical models). Conceptual and theoretical understanding of the OS-PSP relationship is likely to be advanced by drawing upon theories from public administration and management that place OS within a wider organizational ecosystem (e.g. Bidwell and Kasarda, 1985). This would imply conceptualizing relationships as mediated or moderated, rather than independent.

**Implications for research**

The studies examined in this article indicate that when scholars study the OS-PSP relationship they do not take a comprehensive view of the performance construct, in terms of its dimensions and measurement, and perhaps do not fully consider appropriate performance indicators. While the studies reviewed in this article drew upon a range of measures of PSP the majority of studies examined effectiveness (147 samples) as the dependent variable. There was only limited inclusion of measures of efficiency (9 samples), outputs (quality and quantity) (21), responsiveness (29) and indexes (20). It is possible that the slight preponderance of evidence in favor of positive size effects reflects the impact of scale economies on the types of performance measures that are examined in this article. In particular, improvements in effectiveness generally imply increased output levels.

Nevertheless, given the focus of attention on efficiency in many of the theoretical and policy debates on OS, the limited consideration of this measure of performance is troublesome and scholars should seek to examine the effects of OS on PSP across a range of performance indicators simultaneously to tease out the impact of OS on different dimensions of
performance (see, for instance, Andrews et al., 2006).

In examining the OS-PSP relations we argued that scholars should take account of the ways in which the PSP construct is measured as well as its different dimensions. We conducted sub-analysis on the population size effect for perceptual and archival measures of performance. Studies using perceptual measures of performance showed a negative association between OS and PSP while archival presented a positive association, though these results dissipated in the meta-regression. The archival measures of performance used were typically drawn from external sources such as school test results, audited performance indicators or indices of organizational achievements. Perceptual measures were on some occasions taken from managers’ viewpoints (Brewer 2005; Carmeli 2006), but more typically from the perspective of citizens. The negative coefficient in the meta-analysis may therefore reflect large organizations’ perceived lack of responsiveness to this group of stakeholders, as suggested by findings from previous studies of citizen satisfaction (e.g. Denters, 2002; Drew and Dollery, 2016).

OS is a complex construct that can be operationalized in a number of way, we considered static and dynamic measures in our analysis. Light (1999) summarized this debate when he argued that changes in organizational arrangement and personnel hides the “true size of government” and pointed to the effects of contracting out on staffing levels. Research on OS that is able to capture staff on a range of employment terms and those who are contracted to work for government is a stream of research worthy of investigation and which may reveal a different relationship with PSP. Alternatively using proxy measures of OS, notably population may overcome these difficulties. Dynamic measures using population change were a positive moderator suggesting the future research should examine whether size per se or changes in size is the more important determinant of PSP.

OS was extensively used as a control in multiple regression models of management
and PSP, 187 of our 226 samples. We were able to identify these controls by searching for studies on PSP and then interrogating these to identify measures of OS, rather than searching titles and abstracts alone. OS is clearly an important control variable in PSP studies and used to rule out other spurious relationships. We have noted the propensity of authors to offer no hypothesis when including OS as a control: this occurred in relation to 106 samples, or 56.7% of cases. Entering control variables into regression models without full consideration of the likely role of that variable on the dependent variable points towards weak scholarship. Scholars are encouraged to clearly hypothesize the anticipated impact of OS on PSP in future studies. Public administration scholars could learn from recommendations on the use of control variables from management studies. For example, Bernerth and Aguinis (2016) recommend developing theoretical arguments to justify the inclusion of control variables, using citations to show how the control has been used in other studies (including reporting results), providing a hypothesis or anticipating the direction of the relationships between the control and dependent variable. If possible, the validity and reliability of the control measure should also be explored. Given the widespread use of OS as a control, the inclusion of such information would provide clearer evidence on the OS-PSP relationship.

Country effects were identified in the meta-regression. Though at a weaker 0.1 percent of statistical significance, the results showed that larger organizations were associated with PSP in European settings. These findings reflected our expectation, that larger organizations are associated with higher PSP because they have the necessary capacity to manage the higher levels of oversight and regulation seen in this region. It is also possible that regulators in Europe have a preference for closely managing and supporting the performance of big organizations due to the greater risks associated with the failure of such organizations (Pollitt, 2006). While the “European” regulatory size effects noted in the meta-analysis carried through to the meta-regression estimates, the competitive effects of smaller organizations in
the US did not, disconfirming the arguments of public choice scholars about the benefits of government fragmentation. The focus of empirical studies on Europe and the U.S.A. begs questions about the relationship between OS and PSP in other jurisdictions and global regions.

**Limitations**

There are a number of limitations to this study which inform future research on this topic, and we have touched on questions of nonlinearity, performance dimensions and the operationalization of OS and PSP above. In addition to this, the number of studies and samples included in our analysis is relatively small and further research could be conducted in associated fields to ascertain if these findings apply to disciplines such as political science and sociology. In particular, the article only contained 5 studies (6 samples) arguing that smaller organizations were likely to be higher performers. This number of studies is too small to draw any hard and fast conclusions on this hypothesis and researchers are encouraged to undertake further research on these public choice inspired arguments. Finally, while we uncovered a weak relationship between OS-PSP, further interrogation of the studies in our analysis is suggestive of publication bias. As the number of studies on this topic grow others are encouraged to replicate this study to identify if our findings are upheld or the inclusion of additional studies results in new conclusions.

**CONCLUSION**

We find that OS is not a major determinant of PSP. This finding strikes at the heart of longstanding debates from about the role of government efficiency and effectiveness. These debates, drawn from theories of economic, political science and public administration, have variously suggested that “bigger is better” or that “small is beautiful”. We noted that it was claimed that empirical results are equivocal, and the findings reported here could support a new proposition: OS is not a critical antecedent of PSP. However, better theorization of the
complex interplay between OS and other internal and external contingencies in studies of PSP published in public administration journals may assist in addressing this concern, notably considering nonlinear and mediated and moderated relationships.

If there is some veracity to our findings there are profound ramifications for the policy and practice of public administration in respect to the size of government. Governments that make changes to OS may often evoke the prospect of efficiency gains, however the data included in this article would suggest that these claims are rhetorical and imply that changes in OS are motivated by a range of other factors. When governments seek to change the size of organizations they should consider the full spectrum of reasons that are likely to produce positive outcomes. This meta-analysis lays down a research agenda on the OS-PSP hypothesis and we encourage others to contribute further empirical evidence to this debate to advance the theory and practice of public administration.

REFERENCES


Pollitt, Christopher. 2006. Performance management in practice: A comparative study of


Figure 1  Public service performance: Stakeholders and data sources in public service performance with examples of effectiveness in secondary school education

<table>
<thead>
<tr>
<th>Stakeholders</th>
<th>Internal</th>
<th>External</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptual</td>
<td>Staff survey</td>
<td>Citizen survey</td>
</tr>
<tr>
<td></td>
<td>e.g., staff perceptions of</td>
<td>e.g., citizen perceptions</td>
</tr>
<tr>
<td></td>
<td>student attainment</td>
<td>of student attainment</td>
</tr>
<tr>
<td>Archival</td>
<td>Internal data</td>
<td>Audited performance</td>
</tr>
<tr>
<td></td>
<td>e.g., class test results held</td>
<td>indicators</td>
</tr>
<tr>
<td></td>
<td>by teacher</td>
<td>e.g., formal school test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>results</td>
</tr>
</tbody>
</table>

Note: perceptual or archival data can be presented as absolute or relative data.
Figure 2: Symmetric contour funnel plot to identify publication bias
Table 1: Meta-analysis results for the size-performance association

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2$</th>
<th>$r$ [95% CI]</th>
<th>$Z$</th>
<th>$K_s$</th>
<th>$K_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies</td>
<td>727.475***</td>
<td>.006 [-.001, .012]</td>
<td>1.775*</td>
<td>63; 226</td>
<td></td>
</tr>
<tr>
<td>Hypothesized direction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Null</td>
<td>4.833</td>
<td>.000 [-.006, .007]</td>
<td>.107</td>
<td>6; 13</td>
<td></td>
</tr>
<tr>
<td>Larger = higher P</td>
<td>442.098***</td>
<td>.039 [.024, .055]</td>
<td>5.029***</td>
<td>23; 97</td>
<td></td>
</tr>
<tr>
<td>Smaller = higher P</td>
<td>.730</td>
<td>.004 [-.007, .015]</td>
<td>.772</td>
<td>5; 6</td>
<td></td>
</tr>
<tr>
<td>No stated direction</td>
<td>151.246**</td>
<td>-.005 [-.011, .001]</td>
<td>1.770*</td>
<td>35; 110</td>
<td></td>
</tr>
<tr>
<td>Organizational goals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>80.505</td>
<td>.000 [-.003, .003]</td>
<td>.050</td>
<td>23; 87</td>
<td></td>
</tr>
<tr>
<td>Multiple</td>
<td>555.814***</td>
<td>.027 [.014, .039]</td>
<td>4.202***</td>
<td>40; 139</td>
<td></td>
</tr>
<tr>
<td>Size measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>115.169**</td>
<td>.002 [-.006, .010]</td>
<td>.443</td>
<td>27; 80</td>
<td></td>
</tr>
<tr>
<td>Budget</td>
<td>10.190</td>
<td>.008 [-.004, .020]</td>
<td>1.282</td>
<td>18; 37</td>
<td></td>
</tr>
<tr>
<td>School size</td>
<td>8.316</td>
<td>.000 [-.003, .003]</td>
<td>.171</td>
<td>15; 52</td>
<td></td>
</tr>
<tr>
<td>Employees</td>
<td>71.364***</td>
<td>-.010 [-.031, .011]</td>
<td>.962</td>
<td>14; 26</td>
<td></td>
</tr>
<tr>
<td>Population change</td>
<td>50.985*</td>
<td>.148 [.052, .241]</td>
<td>3.022**</td>
<td>7; 31</td>
<td></td>
</tr>
<tr>
<td>Performance dimension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output</td>
<td>8.882</td>
<td>-.002 [-.021, .016]</td>
<td>.240</td>
<td>8; 21</td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>.223</td>
<td>-.002 [-.051, .046]</td>
<td>.093</td>
<td>5; 9</td>
<td></td>
</tr>
<tr>
<td>Effectiveness</td>
<td>679.379***</td>
<td>.006 [-.002, .014]</td>
<td>1.549</td>
<td>41; 147</td>
<td></td>
</tr>
<tr>
<td>Responsiveness</td>
<td>29.729</td>
<td>.005 [-.004, .013]</td>
<td>1.016</td>
<td>9; 29</td>
<td></td>
</tr>
<tr>
<td>Multiple</td>
<td>11.237</td>
<td>.027 [-.006, .061]</td>
<td>1.589</td>
<td>8; 20</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>699.699***</td>
<td>.005 [-.001, .012]</td>
<td>1.547</td>
<td>50; 187</td>
<td></td>
</tr>
<tr>
<td>Performance measurement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptual</td>
<td>41.327</td>
<td>-.011 [-.018, -.004]</td>
<td>3.010**</td>
<td>20; 57</td>
<td></td>
</tr>
<tr>
<td>Archival</td>
<td>702.098***</td>
<td>.007 [.000, .014]</td>
<td>2.062*</td>
<td>49; 169</td>
<td></td>
</tr>
<tr>
<td>Research design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-section</td>
<td>120.880**</td>
<td>-.001 [-.008, .006]</td>
<td>.292</td>
<td>20; 79</td>
<td></td>
</tr>
<tr>
<td>Longitudinal</td>
<td>584.404***</td>
<td>.015 [.005, .024]</td>
<td>2.956**</td>
<td>44; 147</td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>136.646**</td>
<td>-.008 [-.017, .001]</td>
<td>1.782*</td>
<td>35; 89</td>
<td></td>
</tr>
<tr>
<td>EU</td>
<td>554.733***</td>
<td>.016 [.006, .025]</td>
<td>3.307**</td>
<td>23; 113</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>18.082</td>
<td>.016 [.003, .036]</td>
<td>1.615</td>
<td>5; 24</td>
<td></td>
</tr>
</tbody>
</table>

\* $p < .10$, *, p < .05, ** p < .01, *** p < .001
\textsuperscript{a} = Chi-square test of homogeneity of effect sizes
\textsuperscript{b} = r: Population effect size
\textsuperscript{c} = CI: Credibility interval around the population effect size
\textsuperscript{d} = $Z$: Z-score
\textsuperscript{e} = $K_s$: number of studies, $K_c$: number of correlations
Table 2: Meta-regression results for the size-performance association

<table>
<thead>
<tr>
<th>Moderator</th>
<th>β</th>
<th>s.e.</th>
<th>CI95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesized direction (all other are reference category)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larger = Higher P</td>
<td>.005</td>
<td>.012</td>
<td>[-.019, .029]</td>
</tr>
<tr>
<td>No stated direction</td>
<td>-.020*</td>
<td>.011</td>
<td>[-.042, .002]</td>
</tr>
<tr>
<td>Organizational goals (Single is reference category)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple</td>
<td>.011</td>
<td>.010</td>
<td>[-.008, .030]</td>
</tr>
<tr>
<td>Size measure (all other are reference category)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population change</td>
<td>.063***</td>
<td>.016</td>
<td>[.031, .094]</td>
</tr>
<tr>
<td>Performance measurement (Perceptual is reference category)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archival</td>
<td>.009</td>
<td>.010</td>
<td>[-.012, .029]</td>
</tr>
<tr>
<td>Research design (Cross-section is reference category)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longitudinal</td>
<td>.016*</td>
<td>.007</td>
<td>[.001, .031]</td>
</tr>
<tr>
<td>Country (USA is reference category)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU</td>
<td>.016*</td>
<td>.009</td>
<td>[-.001, .033]</td>
</tr>
<tr>
<td>Other</td>
<td>.015</td>
<td>.015</td>
<td>[-.015, .045]</td>
</tr>
<tr>
<td>Constant</td>
<td>-.020</td>
<td>.013</td>
<td>[-.046, .006]</td>
</tr>
<tr>
<td>Chi² (Goodness of fit – QE statistic)</td>
<td></td>
<td></td>
<td>646.413***</td>
</tr>
</tbody>
</table>

p < .10, * p < .05, ** p < .01, *** p < .001