



Quality standards, implementation autonomy, and citizen satisfaction with public services: cross-national evidence

Miyeon Song, Seung-Ho An & Kenneth J. Meier

To cite this article: Miyeon Song, Seung-Ho An & Kenneth J. Meier (2021) Quality standards, implementation autonomy, and citizen satisfaction with public services: cross-national evidence, Public Management Review, 23:6, 906-928, DOI: [10.1080/14719037.2020.1730939](https://doi.org/10.1080/14719037.2020.1730939)

To link to this article: <https://doi.org/10.1080/14719037.2020.1730939>



View supplementary material [↗](#)



Published online: 27 Feb 2020.



Submit your article to this journal [↗](#)



Article views: 460



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)



Quality standards, implementation autonomy, and citizen satisfaction with public services: cross-national evidence

Miyeon Song ^a, Seung-Ho An ^b and Kenneth J. Meier ^{c,d,e}

^aDepartment of Political Science, University of South Carolina, Columbia, SC, USA; ^bSchool of Government & Public Policy, University of Arizona, Tucson, AZ, USA; ^cSchool of Public Affairs, American University, Washington DC, USA; ^dCardiff School of Business, Cardiff University, Cardiff, Wales; ^eInstitute of Public Administration, Leiden University, Leiden, The Netherlands

ABSTRACT

This article investigates whether citizens' evaluations of service performance are related to archival measures of performance, and how institutional context shapes this relationship contingent on administrative autonomy – standards, human resources, and financial autonomy. Using cross-national education data, this study finds that student performance is positively associated with parental evaluations of schools. Perceptions are more closely aligned with performance when agencies have greater autonomy in managing employees, and when national-level bureaucracies set performance standards. This research advances our understanding of the role of administrative autonomy in citizen satisfaction and provides implications for the institutional designs that can benefit performance assessment.

KEYWORDS Citizen satisfaction; performance assessment; bureaucratic autonomy; comparative public administration


Introduction

How citizens view government performance is fundamental for democratic accountability. Citizens' opinions of public services can provide government officials with vital feedback on the job they are doing by alerting officials about changes in service priorities, the clientele they serve, or the need to reallocate scarce resources. Given the importance of citizen inputs for effective democratic governance, governments around the world have gathered citizens' opinions about government performance (Bouckaert, Van de Walle, and Kampen 2005; Van Ryzin 2015) and compared these with administrative performance data. To the extent that citizen evaluations of performance match administrative performance indicators, there is the potential for a consensus on what the government should do, the informed judgements necessary for

CONTACT Seung-Ho An  seunghoan@arizona.edu

Earlier versions of this article were presented at the 76th Midwest Political Science Association Annual Conference, April 5–8, 2018, Chicago, IL, and the 40th Annual Fall Research Conference of the Association for Public Policy and Management, November 8–10, 2018, Washington, DC.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

 The supplementary data for this article can be accessed at <https://doi.org/10.1080/14719037.2020.1730939>

© 2020 Informa UK Limited, trading as Taylor & Francis Group

governmental accountability, and increased citizens' support for overall policy (Hill and Hinton-Andersson 1995; Monroe 1998).

Under the auspices of New Public Management's (NPM) emphasis on performance assessment (Overman 2016, 2017), citizen satisfaction with public services has frequently been used as vital performance information. An implicit assumption underlying the use of citizen satisfaction is that the actual quality of public service matters (Van Ryzin 2015; Van Ryzin, Immerwahr, and Altman 2008) so that there is a strong relationship between citizen satisfaction and government performance. The empirical findings of this relationship, however, are inconclusive (e.g. Favero and Meier 2013; Kelly 2003).

A second assumption is that governments have incentives to be responsive to citizen preferences; therefore, the commonality between citizen evaluations and actual performance exists regardless of institutional contexts. This is partly because most studies have been conducted in decentralized, democratic systems where citizens can signal their preferences to governments through exercising *choice* or *exit* options (Schneider, Teske, and Marschall 2002; Tiebout 1956; but see Brinkerhoff, Wetterberg, and Wibbels 2018; Song and Meier 2018). Under decentralized systems with local control, governments are incentivized to prioritize performance criteria that match citizens' preferences (Tiebout 1956; but see Overman 2017). This assumption, however, does not always hold. Countries with centralized systems or autocratic regimes may have fewer incentives to respond to citizen preferences. In addition, the immature policy environments of developing countries often make it difficult for governments to respond to citizens and manage the performance appraisal system (Nõmm and Randma-Liiv 2012). These raise important questions: Can the association between citizen satisfaction and archival measures of performance be generalized across countries? If so, under what institutional contexts are they most closely aligned?

This article aims to answer these questions and add to the existing literature on citizen satisfaction and public management in three ways. First, this study incorporates variation in institutional contexts across the world and contributes to increasing the generalizability of existing empirical findings. Generalizability is important not only because it is a necessary condition for a good theory, but also because it helps to provide policy implications for performance appraisal and management practice for practitioners who work in different institutional contexts. Although recent studies have broadened the research context by looking at citizen satisfaction in non-US contexts (e.g. Song and Meier 2018; Walker et al. 2018), they are mostly single-country studies and, thus, have a limited ability to consider cross-national variation (but see Brinkerhoff, Wetterberg, and Wibbels 2018). Using a cross-national database, this research advances our understanding of how a country's institutional arrangements make a difference in linking citizen satisfaction to programme performance.

Second, among the various institutional factors, this study incorporates a theoretically and practically important institutional context, *administrative autonomy* – the level of decision-making discretion that administrative agencies have in the policy process – into the research on citizen satisfaction. The delegation of authority over public service delivery to local governments is a global trend allowing more managerial autonomy (Overman 2016; Verhoest et al. 2012); however, we know little about how autonomy affects the link between citizen satisfaction and actual program performance. Examining the role of autonomy therefore makes a significant contribution to the literature.

Third, this research considers both the *degree* of autonomy (how much to delegate) and the *dimensions* of autonomy (what to delegate) and investigates whether citizens' perceptions towards service quality are more closely aligned to archival measures of performance when local agencies have greater autonomy in setting quality standards, managing employees, and allocating financial resources. By doing so, this study provides meaningful information about which dimensions of autonomy should or should not be delegated to agencies. In addition to providing practical implications, this research offers theoretical insights into the multidimensionality of autonomy. As far as we know, our study is among the first that examines how three different dimensions of autonomy affect the relationship between objective and subjective assessments of public services in a cross-national setting.

The empirical context of this research is education, an ideal case to study citizen satisfaction and performance assessment. Quality education with its links to economic mobility is a highly salient issue in most countries, thus providing incentives for parents to be informed about school quality and for public administrators to seek improved performance. Parents are consumers who benefit directly from the service¹; in some cases, parents are permitted some choices in their children's education further increasing the incentives to learn about school performance (Djellal and Gallouj 2009). Unlike other government services where citizens might not use the service and thus be unaware of overall quality, education studies focus on parents with children in school.

We test our research questions using a cross-national education dataset with more than 62,000 individual respondents in 16 countries. The analysis combines archival data on students' academic performance, parents' perceptual judgements of the quality of schools, and administrators' perceptions of autonomy. By doing so, this study investigates under what institutional context parents' perceptual evaluations and archival indicators of school performance are more closely related to each other.

Citizen evaluations and government performance

In response to a growing emphasis on government performance and heightened expectations for quality services, governments seek to measure and evaluate performance accurately. Many government service outcomes, however, are not quantifiable; or there is no consensus among multiple stakeholders on what good performance is. Citizens' perceptual evaluations are often used to measure government performance at both the local and national levels (Bouckaert, Van de Walle, and Kampen 2005), assuming that citizens' assessments are aligned with actual government performance.² Research has sought to investigate whether stakeholders' perceptual judgements and archival performance indicators have some common ground (e.g. Campbell and Fiske 1959; Favero and Meier 2013; Van Ryzin, Immerwahr, and Altman 2008). Archival performance data are quantifiable and observable data on performance (e.g. administrative records of performance), while perceptual data represent stakeholders' perceptual judgements of performance (e.g. citizen satisfaction with service quality) (Andrews, Boyne, and Walker 2006).³

Some scholars argue that citizens' perceptual evaluations may not reflect actual service quality because it is unclear how much accurate information citizens have and what criteria they use to evaluate performance (e.g. Kelly 2003; Stipak 1979). Recent experimental studies also question the validity of citizens' evaluations. They show that citizens unconsciously associate public agencies with inefficiency and inflexibility, thus

biasing their evaluations (Marvel 2016), and that citizen satisfaction does not systematically reflect changes in performance as the theory suggests (Andersen and Hjortskov 2016). Citizens' responses to satisfaction also can be shaped by information cues about government performance (James 2011) and can be contingent on the choice of a positive or negative label description (Olsen 2015). These framing effects raise questions about whether citizens' perceptions of performance accurately represent service quality.

Despite the concern about the validity of citizens' perceptual evaluations, many empirical studies have demonstrated that citizen satisfaction does correlate with service quality measures. In the context of education, Favero and Meier (2013) show that parents' evaluations of New York City schools are positively associated with test scores, objective progress reports, official quality reviews, student attendance, and negatively to school violence (see also Charbonneau and Van Ryzin 2012). In addition to parents' evaluations, research finds that other stakeholders' (students and teachers) perceptual judgements are also significantly related to archival school performance indicators (Song and Meier 2018). Gibbons and Silva (2011), using longitudinal survey data to administrative records on student achievement in England, find a strong relationship between parent satisfaction and academic performance measures. The positive relationship between objective indicators and subjective judgements is not limited to education. Evidence from street cleanliness services also finds a positive correlation between citizen evaluations of service and a cleanliness scorecard (Van Ryzin, Immerwahr, and Altman 2008). Even in a complex policy area with extensive information asymmetry – healthcare, patients' perceptions of service quality are significantly related to objective hospital performance indicators such as clinical process of care scores and 30-day readmission rates (Cheon et al. 2019).

While these studies have made a meaningful contribution, most research linking citizen evaluations to performance indicators has been drawn from Western democratic countries with substantial local autonomy such as the United States, the United Kingdom, and Denmark (Walker et al. 2018). The heavy reliance on a few governance contexts raises concerns about generalizability because how citizens perceive government performance can be influenced by various institutional contexts (Lyons, Lowery, and DeHoog 1992). The underlying theoretical assumptions that local services vary, and that citizens can select jurisdictions that match their tax and services preferences so that citizens' opinions can inform priorities for service provision and influence government funding decisions, however, are not met in many countries (Overman 2017).

As an effort to increase the generalizability of the theory, recent studies have moved the theory to new institutional contexts. Song and Meier (2018) examine how multiple stakeholders' perceptions of school quality are related to archival school performance indicators in South Korea, a centralized education system. Unlike the theoretical works that suggest that the association might be weak in centralized regimes, their study demonstrates a common ground for performance assessment in Korean schools. Walker et al. (2018) also highlight the importance of institutional context and test how perceptual judgements of performance are shaped by performance information in the context of education and solid waste in Hong Kong. In Mexico, Petrovsky, Mok, and León-Cázares (2017) examine whether the theory holds in a developing country with limited accountability and limited competitive democracy (see also Brinkerhoff, Wetterberg, and Wibbels 2018). Even in this context they find that citizens are satisfied

when service quality exceeds their expectations supporting the basic argument of the theory.⁴

Based on these findings, we expect that the positive relationship between citizens' judgements of performance and archival performance indicators will hold in a cross-national context. We test this hypothesis at both the individual-level and the organizational-level and examine whether each level of performance influences citizen satisfaction respectively or jointly. Citizens might evaluate services solely on the basis of how they benefit personally, or they might take a broader view and respond to the benefits to all users. This distinction also provides implications for theoretical works that argue individual benefits matter, but collective ones do not (see Tiebout 1956). We expect that parents are more likely to be satisfied with schools both when their own children achieve higher performance and when the schools perform better.

H1. Parents will give more favourable evaluations to schools when their children perform well on academic achievement tests.

H2. Parents will give more favourable evaluations to schools when the schools perform well on academic achievement tests.

Administrative autonomy and performance evaluations

A trade-off between agency autonomy and democratic accountability remains at the heart of the study of public administration (Kirkhaug and Mikalsen 2009). On the one hand, greater autonomy allows bureaucrats to use their expertise and improve performance (Carpenter 2001). On the other hand, autonomy can be misused to pursue policy goals that diverge from what citizens want (Kogan 2017). Despite the conflict, the devolution of policy responsibilities from higher-level to lower-level governments has been a long-term trend world-wide (Verhoest et al. 2012).

In the present era of delegation, an extensive literature explores the role of public sector agency autonomy (for a review, see Overman 2016). In particular, Overman (2017) explains how agency autonomy shapes citizen satisfaction based on the three theoretical routes. First, according to responsiveness theory, delegation to administrative agencies facilitates managerial discretion in policy implementation, thus, permitting greater responsiveness to citizens and ultimately greater citizen satisfaction with public services (Lyons, Lowery, and DeHoog 1992; Van Thiel 2001). Second, building upon the blame-shifting argument, delegation can be a strategy for politicians and the central government to hide behind agencies when things go wrong (Hood 2002). Autonomous agencies often 'serve as a shield to deflect blame for bad service outcomes,' which could increase citizens' satisfaction with politicians and the central government (Overman 2017, 215). Third, based on credibility theory, the delegation of task implementation can be seen as a commitment to independent decision making, and such depoliticization can improve the impartiality of policy implementation (Knott and Miller 2006) as well as continuity in service delivery (Overman 2017). These will lead to higher citizen satisfaction.

Although Overman (2017) focuses on citizens' satisfaction with the central government rather than service agencies, the logic should apply to our study. First, delegation to schools facilitates a principal's discretion in school policy and educational curriculum and allows greater responsiveness to parents and students. When schools offer

more tailored education, parents should be more satisfied with the schools. Second, when schools have greater autonomy, parents should be more able to blame schools for bad performance than when the central government has complete control over education policy (Koppell 2005).⁵ Third, administrative autonomy helps protect schools from politically-driven education policies (see Hammond et al. 2019) and allows schools to continue to offer stable and sustainable education programmes. This can lead to higher parent satisfaction with schools.

In practice, administrative autonomy varies significantly across countries as well as within countries. This variation should affect not only parent satisfaction but also the relationship between parent satisfaction and archival school performance indexes. Some countries, such as the United States, have decentralized education systems that grant schools significant autonomy. Parents in such systems can send preference signals to schools by exercising school choice options or participating in school decision making (Schneider, Teske, and Marschall 2002). Under this context, parents are more likely to be happy with schools, and administrative school performance indicators should be more closely related to parents' satisfaction with their children's schools (see Favero and Meier 2013). Other countries have centralized systems that limit school autonomy and parents' school choice. In this context, it is difficult for parents to express their preferences and needs; therefore, their satisfaction would not closely be linked to objective performance indicators set by the central government. Little empirical evidence, however, exists on this question.

Another important gap in the literature is the multidimensionality of autonomy. Previous literature has focused on the degree of autonomy and paid less attention to the question of how different dimensions of autonomy affect public service delivery. Recognizing the multidimensionality of autonomy (Krause and Van Thiel 2019; Verhoest et al. 2012; Wynen et al. 2014), we argue that '*what (not) to delegate*' matters as much as '*how much to delegate*' in shaping the relationship between subjective and objective performance assessments. In particular, this study considers three distinct aspects of autonomy – (1) standards, (2) human resources, and (3) financial autonomy. Standards autonomy reflects devolving authority to administrative agencies to set the actual quality standards used in evaluation whereas human resources and financial autonomy capture delegating the responsibility of the agencies to get the job done in the policy implementation process. Based on the literature suggesting that the degree and the type of autonomy allowed to administrative agencies varies within and between countries (Overman 2017; Verhoest et al. 2012), we expect that each type of autonomy can have a unique moderating effect on the relationship between archival performance indicators and citizens' perceptual judgements of service quality.

Standards, human resources, and financial autonomy

Standards autonomy

Standards autonomy is defined as the degree of decision-making authority in setting quality standards for public programs. The delegation of authority to set standards directly relates to an autonomy-accountability dilemma, because if agencies have significant autonomy in the early stages of policy goal setting (or quality standards setting), it is more difficult for elected officials to hold agencies accountable (Nielsen 2014). Standards autonomy also involves the inherent risk of goal displacement with

individual agencies defining their own policy objectives. Public agencies may have incentives to focus on tangible policy *outputs* rather than more meaningful policy *outcomes* to maximize their performance ratings (Merton 1940), engage in ‘effort substitution (reducing effort on nonmeasured performance dimensions) or gaming (making performance on the measured performance dimension appear better, when in fact it is not)’ (Kelman and Friedman 2009, 918).

Greater standards autonomy may lead agencies to seek lower standards because it is easier for them to exceed expectations and get high performance ratings when standards are low. High standards that are difficult to meet may not be favourable to the agency even though the pursuit of higher service quality is good for citizens. In education, permitting states to set performance standards under the US *No Child Left Behind* Act allowed low performing states to set significantly lower standards (see Manna 2006). In addition, individual schools may try to manipulate student performance on standardized exams by purposely excluding low achievers from the testing (see Bohte and Meier 2000). As Bohte and Meier (2000, 180) note, ‘cheating is likely to occur in organizations in which the day-to-day activities of bureaucrats are not heavily monitored (for example, highly *decentralized* bureaucracies).’

Based on the discussion, we expect that greater local autonomy in setting quality standards can negatively affect the link between archival performance indicators and citizen evaluations. When schools have greater autonomy in setting assessment policies and educational curricula, they might set low standards (e.g. adopting lenient performance standards, lenient grading, or offering easier courses) to maximize their performance ratings. The performance indicators schools use, in this case, may not capture what parents value, thus increasing the gap between archival performance measures and parents’ evaluations of school quality. Under standards set by national governments, in contrast, schools have less space for manipulating the quality standards; therefore, administrative performance indicators may more closely reflect citizens’ judgements of school performance.

When national quality standards exist, administrative performance can be compared across different schools; knowing how their children’s schools perform compared to other schools may help parents make more accurate judgements about school quality. Highlighting the notion of social aspirations (Cyert and March 1963) where organizations or individuals compare themselves to others, Barrows et al. (2016) argue that parents’ evaluations of schools can be influenced by how their child’s school ranks relative to other schools at the state, national, or international levels. Decentralized and fragmented systems with different quality standards, therefore, make it more difficult for parents to evaluate the quality of their school compared to other schools (see also Olsen 2017).

H3a. Student performance will be less closely aligned to parents’ evaluations in schools with greater standards autonomy.

Human resources autonomy

Human resources autonomy is defined as the degree of decision-making authority in managing people. It concerns the devolution of responsibilities to administrative agencies to hire, fire, compensate, train, and motivate employees. The authority to manage human resources is especially important in labour-intensive policy areas

because effective personnel management directly relates to better service quality and increased citizen satisfaction (Favero, Meier, and O'Toole 2016). Greater human resources autonomy can help agencies attract and hire the employees who fit the job or bring local cultural skills, thus allowing the organization to provide better service (Nielsen 2014).

Responsiveness theory presumes that the delegation of autonomy facilitates interaction between bureaucrats and citizens allowing for more locally tailored service (Van Thiel 2001) that leads to increased citizen satisfaction (Overman 2017).⁶ Greater human resources autonomy allows schools to hire teachers and staff who can provide more customized service to students (e.g. teachers specialized in teaching nonnative speakers). In a similar vein, the decentralization literature argues that the devolution of power from upper-level to lower-level governments can make bureaucracies more responsive and accountable to citizens (Escobar-Lemmon and Ross 2014). In theory, fragmented and decentralized local governments can provide a greater range of public services responding to the preferences of citizens, whereas centralized national governments often implement *one size fits all* policies (Li, Wang, and Zheng 2017).⁷ Human resources autonomy can also encourage *co-production*, therefore, contributing to reducing the gap between citizens' perceptions of services and administrative performance indicators (Jakobsen 2012).

The education literature has also suggested that school autonomy encourages parental involvement and co-production (Bifulco and Ladd 2005). Increased parental involvement can contribute not only to school quality or student performance (e.g. Marschall 2006) but also can reduce the gap between parents' perceptions of school quality and archival performance indicators. This is possible because the more parents participate, the more their opinions are likely to be reflected in educating students and the more they know what the school is actually doing.

H3b. Student performance will be more closely aligned to parents' evaluations in schools with greater human resources autonomy.

Financial autonomy

Financial autonomy is defined as the degree of decision-making authority in financial transactions (or authority in determining revenues and expenses). Similar to human resources autonomy, financial autonomy can also help agencies to provide more customized services to citizens and encourage citizen engagement by allowing agencies to allocate financial resources according to citizens' preferences and needs.

While the literature on fiscal decentralization supports the argument for providing more localized services, the participatory budgeting scholarship suggests two competing hypotheses about the effect of financial autonomy on citizen participation. The first scenario proposes that managers in an agency with greater financial autonomy would rely on their own expertise and knowledge to make budget decisions rather than seeking citizens' opinions. An alternative perspective holds that greater autonomy in budgeting leads to accountability, and managers are more likely to engage citizens to increase the legitimacy of agency actions (Neshkova 2014). Empirical findings also show mixed results. Neshkova (2014), for instance, finds that greater autonomy in allotment processes and own-source revenue has a positive relationship with citizen

participation, whereas greater autonomy in developing spending forecasts is negatively correlated with citizen participation in US transportation and environmental policy.

When a school has a greater autonomy in budget formulation and allocating funds, parents might feel that they can exert greater influence over educational programs and policies and gain more from participating; therefore, they may be more willing to devote their time and effort to participate (Bifulco and Ladd 2005, 556; Parrado et al. 2013). Put differently, greater financial autonomy can make involvement worthwhile for parents by providing parents the material incentives to participate in collective efforts to influence school programs and policy. When parents engage with their children's schools, their evaluations of school quality are more likely to match administrative school performance indicators because parents understand the production process.

H3c. Student performance will be more closely aligned to parents' evaluations in schools with greater financial autonomy.

Data and methods

Testing these hypotheses requires cross-nationally comparable data on parents' perceptions of school quality, an archival measure of school performance, and an autonomy index. We use the Organization for Economic Cooperation and Development's (OECD 2015) Program for International Student Assessment (PISA) database. PISA provides measures of 15-year old students' academic performance in maths, science, and reading, and this measure allows countries to compare student learning outcomes. In 2015, more than 70 countries and a few education systems that are not countries (e.g. Hong Kong and Macao) participated in the PISA assessment.

Within participating countries, samples were drawn via a multistage stratified random sampling process designed to generate a fully representative sample for each country (OECD 2015).⁸ While 72 countries and education systems participated in the academic performance assessment, only 16 countries participated in the parent survey.⁹ Our sample includes Chile, Croatia, Dominican Republic, France, Georgia, Germany, Hong Kong, Ireland, Italy, Korea, Luxembourg, Macao, Malta, Mexico, Portugal, and Spain. While these countries are clearly not representative of all nations of the world, they do provide a wide range of countries that can provide a general test of the hypotheses.¹⁰

Since our data include both individual and organization level variables, ignoring the multilevel data structure can bias standard errors downward. Since we pool countries together in estimating our results, the country level characteristics should also be considered in the modelling strategy. One way to address this issue is to employ three-level multilevel models with country level control variables. A challenge with multilevel modelling is that the group N should be large enough (otherwise, inefficient; see the discussion from Maas and Hox 2005); the small degrees of freedom at the country level also limits us accounting for various country characteristics. For these reasons, rather than using multilevel models, we employ a mixed modelling strategy.¹¹ We use Ordinary Least Square (OLS) regression models with clustered standard errors by schools accounting for the first two levels. In addition, to control for unobserved country-specific effects on parents' satisfaction with schools, all models include country fixed effects. Lastly, we also use the

sampling weight of grade non-response adjustment in each country provided by PISA.

Measures

Dependent variable: perceived school quality

The dependent variable of theoretical interest is perceived school performance as measured by parents' perceptual judgements of the schools. The PISA parent survey asks about various aspects of the school quality, such as teacher quality, education standards, teaching methods, disciplinary atmosphere, and the quality of education, on a four-point scale (from strongly disagree to strongly agree). Factor analysis demonstrates that all survey items loaded onto a single factor producing an eigenvalue of 4.00 with a Cronbach's alpha of 0.87. [Table A1](#) in the Appendix shows the details of the survey questions and the factor analytic results.

Independent variables: archival performance indicators

This study adopts two different levels of archival performance indicators as independent variables – individual student test scores and school mean scores. All students in the sample take tests in maths, science, and reading; and we create two standardized performance indexes using these test scores (one for individual scores, the other for school scores). This permits us to determine if parents evaluate schools based on their own child's performance or the performance of the entire school. Although PISA scores are not an official performance indicator in any of the countries examined, they are highly correlated with other standardized exams that are used in official assessments (Rindermann 2007).

Standards, human resources, and financial autonomy

The article incorporates three dimensions of autonomy: standards, human resources, and financial autonomy. We measure each dimension of autonomy using questions about who has the responsibility for educational standards, staffing, and budgeting. The PISA school survey includes a set of questions, 'regarding your school, who has considerable responsibility for the following tasks?' School administrators report whether (1) teachers, (2) the principal, (3) the school governing board, (4) the regional or local education authority, or (5) the national education authority has a responsibility for the tasks such as hiring and firing teachers, determining teacher salaries, formulating and allocating school budgets, establishing disciplinary policies and assessment policies, determining course content and textbooks, etc. For each item, we assign a value of four if inside school actors (the principal or teachers) have the responsibility, three for the school governing board, two for the regional/local education authority, and one for the national education authority.¹² The underlying logic here is that the more upper-level agencies (education authorities) delegate authority for managing schools to school-level actors (principals and teachers), the more autonomy schools have.

The factor analysis of school autonomy indicators produces a three-factor solution (see [Table A2](#) in the Appendix).¹³ The first factor taps school autonomy in assessment

policies and educational curricula reflecting who has decision making power in setting educational standards. We use this factor to measure standards autonomy. The second factor mainly captures school autonomy in managing human resources (teachers). The third factor captures autonomy in formulating and allocating the school budget, and we use this factor as financial autonomy.

Control variables

The citizen satisfaction literature suggests that socioeconomic and demographic factors are significant predictors of citizens' preferences and satisfaction (Brown and Benedict 2002). We control for the parents' educational attainment (coded as a six category-variable from no formal education to tertiary education and advanced research programs) and the students' immigrant status (native = 0; second generation = 1; first generation = 2) at the individual level. We also consider whether a student has repeated a grade or not (grade repetition = 1; otherwise = 0). At the school level, the ownership of the school (public = 1; private = 0) and the percent government funding (ranges from 0% to 100%) that captures financial publicness are controlled. School size is measured by the number of students at the school (logged). The models also include school resources such as the percentage of full-time teachers (ranges from 0% to 100%) and class size (the student-teacher ratio). Table A3 in Appendix shows descriptive statistics and coding schemes for all variables.

Findings

The objectives of the analysis are to test (1) whether the relationship between perceptual evaluations (parents' satisfaction with schools) and archival performance measures (student and school performance) is generalizable and (2) how standards, human resources, and financial autonomy shape the relationship between the two. Before examining our main research questions, we explore the cross-national variation in the link between archival and perceptual measures. Table 1 shows the regression coefficients and t-scores for each country. All control variables are included in all models. The effect of academic performance on parents' evaluations of their children's school (both the direction and the significance of the effect) varies substantially across countries. In some countries, both student-level and school-level performance have a positive and significant relationship with parents' satisfaction (e.g. Hong Kong and Mexico). In other countries, however, only either student-level performance (Dominican Republic and Malta) or school-level performance (Chile, Croatia, Italy, Korea, and Portugal) has a positive association with parents' satisfaction.

The relatively modest size of the relationships needs to be interpreted within the context of performance assessment in education. The PISA scores are not an official government-sanctioned measure of education performance in any of these countries. Countries may have other official test indicators or no indicators at all. Similarly, countries such as Germany do not report any test scores to parents, and parents would not know the PISA scores of their child. Although PISA scores are positively correlated with other standardized tests including those used for national assessments (Rindermann 2007), the correlation between perceptions of school performance and actual test scores relies on a series of judgements on the part of parents in an environment that may not be information rich. Despite these difficulties, we find a

Table 1. Cross-national variation in the relationship between archival performance indicators and perceptual evaluations.

DV = Parents' evaluations	IV = Student performance		IV = School performance		N
	Coef.	T-scores	Coef.	T-scores	
Overall	0.04**	2.59	0.06**	2.66	62,916
Chile	0.00	0.02	0.23**	3.08	4,726
Croatia	-0.04	-1.64	0.12**	3.88	4,737
Dominican Republic	0.16**	2.90	0.08	1.15	2,862
France	0.03	0.83	0.02	0.52	3,953
Georgia	-0.04	-1.52	0.05	0.90	2,965
Germany	0.00	-0.10	-0.11+	-1.87	2,029
Hong Kong	0.07**	3.29	0.07+	1.92	4,167
Ireland	0.02	0.61	0.00	-0.03	3,693
Italy	-0.03	-0.73	0.15**	4.12	5,358
Korea	0.01	0.34	0.21**	4.23	4,998
Luxembourg	0.02	0.84	0.08	1.18	2,691
Macao	0.03	1.54	0.01	0.28	4,251
Malta	0.17**	4.78	-0.08	-0.52	2,109
Mexico	0.10*	2.19	0.11+	1.84	4,916
Portugal	-0.04	-1.09	0.07*	1.96	5,472
Spain	0.02	0.50	0.02	0.33	3,989

All controls included not shown; clustered robust standard errors by school; + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; two-tailed tests.

relationship between actual test scores and parents' perceptions of performance. Because this relationship varies greatly across the nations in our study, the next step is to determine what factors can enhance the ability of clients (parents) to evaluate the quality of services (education).

Table 2 shows the pooled regression models to test the generality of theory. Based on the theory and evidence from previous single country studies that offer a positive association between parents' evaluations and student performance, we first hypothesize that the positive link can be generalizable across countries. Consistent with our theoretical expectations, Model 1 in Table 2 shows both individual student performance (test scores) and overall school performance are positively and significantly associated with parents' evaluations of school quality (H1 and H2 supported). Parents are more satisfied with their child's school when the school achieves higher academic performance as well as when their own children do well on tests. The positive relationships remain the same when we add the three autonomy measures (Model 2). The interaction of student and school performance, however, is unrelated to parent satisfaction (Model 3).

The primary research interest of this study is to examine the role of autonomy in shaping the relationship between perceptual evaluations and archival assessments. By adding interaction terms between archival indicators of performance and the three autonomy measures, we test whether and how autonomy moderates the effect of archival assessments on satisfaction (Table 3). The interaction between student performance and standards autonomy shows a negative and significant coefficient suggesting that when schools have more autonomy in setting standards, student performance is less associated with parents' satisfaction (Model 1 in Table 3; H3a supported). By contrast, human resources autonomy positively moderates the effect of student performance on parents' evaluations, indicating that parents' judgements of school quality and student performance are more closely aligned when schools have more responsibilities in

Table 2. A general model for the relationship between archival performance indicators and perceptual evaluations.

Dependent variable = Parents' evaluations			
	Model 1	Model 2	Model 3
Student performance	0.041** (0.015)	0.040** (0.015)	0.041** (0.015)
School performance	0.054* (0.022)	0.056** (0.021)	0.057** (0.021)
Standard autonomy		0.057+ (0.031)	0.058+ (0.031)
Human resources autonomy		−0.004 (0.026)	−0.003 (0.026)
Financial autonomy		0.005 (0.023)	0.006 (0.023)
Student performance × School performance			0.004 (0.012)
Highest parent education	−0.037** (0.008)	−0.038** (0.008)	−0.038** (0.008)
Immigrant status	0.010 (0.023)	0.009 (0.022)	0.009 (0.022)
Grade repetition (yes = 1; no = 0)	−0.066+ (0.035)	−0.076* (0.035)	−0.077* (0.035)
Ownership (public = 1; private = 0)	−0.228** (0.040)	−0.230** (0.051)	−0.230** (0.051)
Government funding (%)	−0.002* (0.001)	−0.001+ (0.001)	−0.001+ (0.001)
Full-time teachers (%)	0.000 (0.001)	0.001 (0.001)	0.001 (0.001)
Number of students (logged)	−0.051* (0.024)	−0.051* (0.024)	−0.050* (0.024)
Student-teacher ratio	0.007** (0.002)	0.007** (0.002)	0.007** (0.002)
Constant	0.531** (0.161)	0.498** (0.170)	0.495** (0.169)
R-Squared overall	0.069	0.070	0.070
N	62,916	62,916	62,916

Clustered robust standard errors by school in parenthesis; country fixed effects included but not shown; + p < 0.10, *p < 0.05, **p < 0.01; two-tailed tests.

managing teachers (Model 2; H3b supported). The interaction between financial autonomy and student performance is not statistically significant (Model 3; H3c not supported). These results remain the same in the full model (Model 4).

A more intuitive way of illustrating these results is plotting predicted values. Figure 1 shows predicted parents' evaluations in a school at varying levels of student performance given different levels of standards autonomy. The solid line shows the relationship between student performance and parents' evaluations when schools have a high level of standards autonomy (two standard deviations above the mean), and the dashed line illustrates the relationship for a low level of standard autonomy (two standard deviations below the mean). The slope of the solid line is negative, suggesting that student performance is negatively associated with parents' evaluations when schools have more standards autonomy relative to upper-level bureaucracies. The slope of the dashed line is positive, by contrast, indicating that student performance is positively related to parents' evaluations when schools have less standards autonomy.

Table 3. The role of autonomy in the relationship between archival performance indicators and perceptual evaluations: individual level performance.

Dependent variable = Parents' evaluations				
	Model 1	Model 2	Model 3	Model 4
Student performance	0.053** (0.017)	0.054** (0.016)	0.040** (0.015)	0.062** (0.017)
School performance	0.058** (0.021)	0.059** (0.021)	0.056** (0.021)	0.060** (0.021)
Standard autonomy	0.029 (0.027)	0.060+ (0.031)	0.058+ (0.030)	0.035 (0.026)
Human resources (HR) autonomy	-0.011 (0.025)	0.017 (0.026)	-0.003 (0.026)	0.006 (0.026)
Financial autonomy	0.001 (0.021)	0.004 (0.023)	0.007 (0.018)	-0.000 (0.017)
Student performance × Standard autonomy	-0.045** (0.017)			-0.038* (0.018)
Student performance × HR autonomy		0.037** (0.013)		0.028* (0.014)
Student performance × Financial autonomy			0.003 (0.019)	-0.001 (0.018)
Highest parent education	-0.039** (0.009)	-0.039** (0.008)	-0.038** (0.008)	-0.039** (0.009)
Immigrant status	0.010 (0.022)	0.007 (0.022)	0.009 (0.023)	0.008 (0.023)
Grade repetition (yes = 1; no = 0)	-0.069* (0.035)	-0.080* (0.035)	-0.076* (0.035)	-0.073* (0.035)
Ownership (public = 1; private = 0)	-0.235** (0.050)	-0.217** (0.051)	-0.229** (0.051)	-0.225** (0.051)
Government funding (%)	-0.001+ (0.001)	-0.001+ (0.001)	-0.001+ (0.001)	-0.001+ (0.001)
Full-time teachers (%)	0.001 (0.001)	0.000 (0.001)	0.001 (0.001)	0.001 (0.001)
Number of students (logged)	-0.052* (0.025)	-0.050* (0.024)	-0.050* (0.024)	-0.052* (0.024)
Student-teacher ratio	0.007** (0.002)	0.007** (0.002)	0.007** (0.002)	0.007** (0.002)
Constant	0.503** (0.170)	0.490** (0.169)	0.496** (0.168)	0.497** (0.168)
R-Squared overall	0.071	0.071	0.070	0.072
N	62,916	62,916	62,916	62,916

Clustered robust standard errors by school in parenthesis; country fixed effects included but not shown; + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$; two-tailed tests.

Figure 2 illustrates the predicted effects of student performance on parents' evaluations given different levels of human resources autonomy. The plot shows the opposite pattern from Figure 1, with a positive relationship between student performance and parents' evaluations for schools with a high level of human resources autonomy (the solid line) and a slightly negative relationship with a low level of human resources autonomy (the dashed line).¹⁴

Recognizing individual student level performance and aggregate school-level performance may have different effects on parents' satisfaction, we conduct additional analyses testing the interactive relationship between autonomy and academic performance at the school level (Table 4). The findings from the school level performance are similar to those from the student level performance model. Whether the performance criterion is the entire school or the individual child, parents' perceptions of quality education are more consistent with these measures when local schools have substantial autonomy in human resources but limited autonomy in setting performance standards

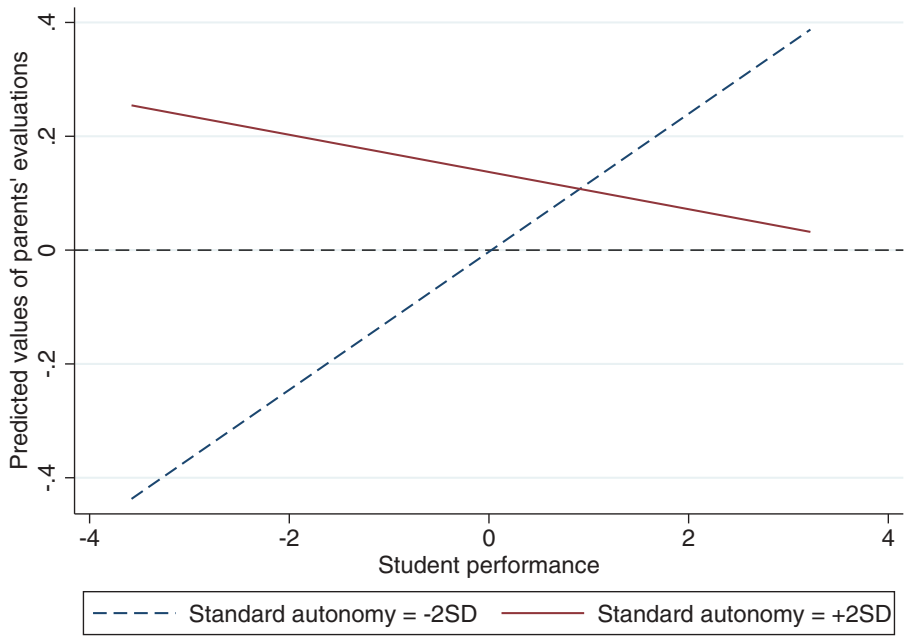


Figure 1. Predicted effects of student performance on parents' evaluations depending on the levels of standard autonomy.

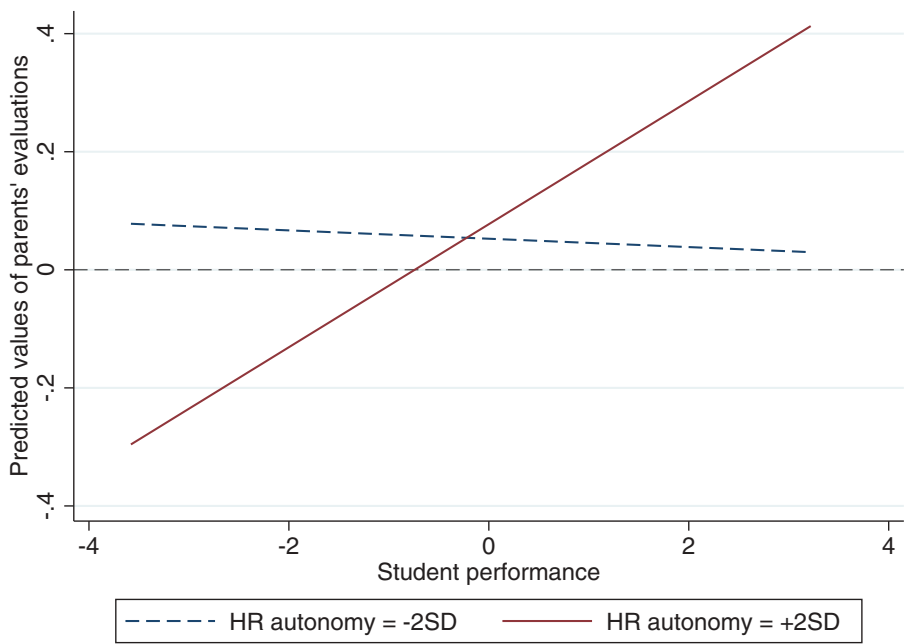


Figure 2. Predicted effects of student performance on parents' evaluations depending on the levels of human resources autonomy.

Table 4. The role of autonomy in the relationship between archival performance indicators and perceptual evaluations: aggregate level performance.

Dependent variable = Parents' evaluations				
	Model 1	Model 2	Model 3	Model 4
Student performance	0.040** (0.015)	0.039* (0.015)	0.040** (0.015)	0.040* (0.015)
School performance	0.068** (0.022)	0.076** (0.023)	0.057** (0.021)	0.083** (0.024)
Standard autonomy	0.025 (0.025)	0.061* (0.031)	0.060* (0.030)	0.035 (0.025)
Human resources (HR) autonomy	-0.012 (0.025)	0.025 (0.027)	-0.002 (0.026)	0.013 (0.026)
Financial autonomy	0.000 (0.021)	0.003 (0.023)	0.019 (0.017)	0.012 (0.017)
School performance × Standard autonomy	-0.039* (0.019)			-0.033 (0.020)
School performance × HR autonomy		0.038** (0.014)		0.030* (0.014)
School performance × Financial autonomy			0.016 (0.020)	0.015 (0.019)
Highest parent education	-0.039** (0.009)	-0.039** (0.008)	-0.038** (0.008)	-0.039** (0.009)
Immigrant status	0.008 (0.022)	0.008 (0.022)	0.009 (0.023)	0.007 (0.023)
Grade repetition (yes = 1; no = 0)	-0.072* (0.035)	-0.082* (0.035)	-0.075* (0.036)	-0.076* (0.035)
Ownership (public = 1; private = 0)	-0.237** (0.050)	-0.212** (0.052)	-0.221** (0.052)	-0.213** (0.051)
Government funding (%)	-0.001* (0.001)	-0.001+ (0.001)	-0.001+ (0.001)	-0.001+ (0.001)
Full-time teachers (%)	0.001 (0.001)	0.000 (0.001)	0.001 (0.001)	0.001 (0.001)
Number of students (logged)	-0.052* (0.025)	-0.050* (0.024)	-0.050* (0.024)	-0.051* (0.024)
Student-teacher ratio	0.007** (0.002)	0.007** (0.002)	0.007** (0.002)	0.007** (0.002)
Constant	0.505** (0.170)	0.485** (0.169)	0.482** (0.167)	0.479** (0.167)
R-Squared overall	0.071	0.071	0.070	0.072
N	62,916	62,916	62,916	62,916

Clustered robust standard errors by school in parenthesis; country fixed effects included but not shown; + p < 0.10, *p < 0.05, **p < 0.01; two-tailed tests.

for the organization. Fiscal autonomy appears to have little impact on the congruence of citizen perceptions and organizational performance.¹⁵

Discussion and conclusion

This article advances our understanding of how citizens' perceptual judgements of public services relate to archival indicators of service quality across countries and how institutional structures affect this relationship. By incorporating three dimensions of autonomy (standards, human resources, and financial autonomy), this study provides theoretical clarity and practical implications for institutional designs that can facilitate government performance assessment. Based on the literature on goal displacement, decentralization, and co-production, we hypothesized that standards autonomy

negatively moderates the relationship, whereas human resources and financial autonomy positively moderate the link.

Analyses from more than 62,000 parents in 16 different countries suggest that parents' perceptual evaluations of schools are significantly associated with student test scores. This result implies that the convergent validity of the performance indicators can be generalized, although there are exceptions to this pattern. The finding also indicates that countries can design performance systems to enhance the ability of parents to evaluate schools. Parents' satisfaction and student performance are more closely aligned when schools have greater autonomy than national-level education authorities in managing teachers. Interestingly, standards autonomy shows an opposite pattern, suggesting that student performance has a stronger relationship with parents' satisfaction when a national-level bureaucracy has greater authority in setting the standards in education.

The opposite effect of standard and human resources autonomy is both theoretically and practically meaningful in understanding and improving citizens' satisfaction with public services. Theoretically, our finding highlights the multidimensional nature of autonomy (Verhoest et al. 2012) by demonstrating each dimension of autonomy has a unique role in shaping the relationship between citizens' perceptual evaluations and archival performance. This result is consistent with a Danish study that shows that school autonomy (defined as a managerial authority) over human resources positively moderates the effect of performance management while school autonomy in goal setting has a negative moderating effect (see Nielsen 2014). Also, the findings of our study suggest that the convergence of performance indicators can be conditional on an institutional context rather than being absolute. The crucial institutional context that contributes to convergent validity (at least in the context of education) is local flexibility in managing street-level bureaucrats while subject to consistent national standards.

This article provides practical implications by answering the questions of what to delegate and what not to delegate. The delegation of authority and responsibility became more popular under NPM (Verhoest et al. 2012) based on the assumption that greater managerial autonomy can increase service user satisfaction as well as overall performance. Our findings suggest that the devolution of autonomy is not a panacea. In particular, decentralized standards setting might encourage managers to set low standards of performance with detrimental consequences for desirable policy outcomes (see also Nielsen 2014). Variation in standards across jurisdictions also muddies the signal to citizens about the absolute quality of public services. This finding suggests that quality standards setting should not be left to local administrative organizations or managers, but set centrally to create greater transparency and accountability. Empowering local authorities over human resources, by contrast, can contribute to more positive perceptions of government among citizens and greater convergent validity of the performance indicators. In sum, the institutional design that allows more flexibility in people management (delegating human resources management) with centrally set standards (not delegating the quality standard setting) can contribute to a better performance appraisal system.

There are several limitations of this study, which inform future research on this topic. While this research highlights the value of cross-national investigation, our empirical models do not necessarily explore the role of national-level factors. Future research should examine how macro-level structures (e.g. democratic vs. autocratic

regimes, political decentralization) shape the link between citizens' subjective judgments and objective performance indicators. Whether national-level structures interact with within-country autonomy when shaping citizen satisfaction with governments also merits study. Citizen satisfaction with various levels of government also provides the potential for future research, given that the delegation of decision-making authority involves the power dynamic between local and national governments. While this study focuses on citizen satisfaction with service organizations at the local level, future research can further explore citizen satisfaction with national government with varying level of administrative autonomy.

Lastly, it is worth discussing the question of whether our findings from education could be transferable to other public services such as healthcare, transportation, or welfare. In education, parents and students benefit directly from the education services and are likely to be aware of the quality of schools. In many other government services, however, citizens benefit indirectly from those services and are less likely to aware of overall quality. Although the current study only dealt with schools, we expect that the logic of national standards and local flexibility on human resources to be applicable to other public services. National standards give clients a uniform standard and provide clarity that facilitates comparative evaluation. Local flexibility in human resources permits crafting policies to fit local needs. While both factors are likely to enhance the client's ability to evaluate a wide range of services, only additional studies in other areas can indicate how general this finding is.

Notes

1. There are, of course, other service areas that citizens are the direct consumers, such as utilities, waste removal, public transportation, fire protection, etc.
2. The expectations-disconfirmation model (EDM) highlights the role of citizens' expectations as well as the actual quality in shaping satisfaction with services (e.g. James and Moseley 2014; Morgeson 2013; Van Ryzin 2004). The theory suggests that citizen satisfaction is shaped by the discrepancy between prior expectations and experienced performance (James and Moseley 2014; Van Ryzin 2004).
3. Both measures have their own pros and cons. Archival performance indicators have been regarded as desirable due to being independent from perceptual judgements (Andrews, Boyne, and Walker 2006); however, they often only focus on service aspects that can be easily quantified and may not represent actual service quality. Perceptual performance measures, in contrast, can capture non-quantifiable but important elements of performance that matter to citizens, but are limited by citizen knowledge about the service.
4. These studies, however, do not necessarily seek to account for national contexts in explaining citizen satisfaction and are not comparative in nature.
5. Parents are more likely to have some sense of 'who makes the decision in schools' because they often have opportunities to observe the school decision making process via joining Parent Teacher Associations (PTA), participating in school events, or attending school meetings or board of education meetings.
6. Unlike the theoretical expectation, Overman (2017), finds no statistical differences in citizen satisfaction with public services between autonomous agencies and government units.
7. Decentralized systems at the same time might not be as vigilant at protecting minority rights.
8. For more details, see OECD (2015).
9. A small but unrepresentative sample of parents in the United Kingdom also participated in the parent survey and were excluded from the analysis. The substantive results in this paper remain the same whether or not we include the UK data.
10. The Chow test results show that pooling data is essential instead of estimating the results in each country (results available upon request).

11. As a robustness check, we also run our models with the three-level multilevel modelling approach and find that our results remain largely the same (results available upon request).
12. A school administrator can mark a place for more than one actor. In this case, we calculate the mean of the assigned values. For instance, if a principal checks boxes for the school governing board (3) and the regional/local education authority (2), we assign a value of 2.5.
13. To create the three autonomy factors, we use the entire sample in PISA 2015 rather than the sixteen countries that are used in the regression models, since the two resource factors did not distinctively load onto two separate factors in the smaller sample. The substantive results, however, remain the same whether we use the three autonomy factors via the entire sample in PISA 2015 or via the ones used in the smaller sample.
14. The plots of the marginal effects are also available in the online appendix.
15. The plots of marginal and predicted effects from Table 4 are available in the online appendix.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Miyeon Song (misong@sc.edu) is an assistant professor in the Department of Political Science at the University of South Carolina. Her research focuses on public management, organizational theory and behaviour, government performance, citizen satisfaction, and representative bureaucracy.

Seung-Ho An (seunghoan@arizona.edu) is an assistant professor in the School of Government and Public Policy at the University of Arizona. His research interests include public and non-profit management, representation, bureaucratic politics, organizational behaviour, leadership, and research methods.

Kenneth J. Meier (kmeier@american.edu) is a distinguished scholar in residence in the School of Public Affairs at American University and holds faculty appointments at the Cardiff School of Business (UK) and Leiden University (the Netherlands). His research interests include public management, equity and public policy, methodological issues in the social sciences, and new theories of representative bureaucracy, organizational environments, and managerial decision making.

ORCID

Miyeon Song  <http://orcid.org/0000-0002-2875-9786>

Seung-Ho An  <http://orcid.org/0000-0002-5998-2411>

Kenneth J. Meier  <http://orcid.org/0000-0002-6378-0855>

References

- Andersen, S. C., and M. Hjortskov. 2016. "Cognitive Biases in Performance Evaluations." *Journal of Public Administration Research and Theory* 26 (4): 647–662. doi:10.1093/jopart/muv036.
- Andrews, R., G. A. Boyne, and R. M. Walker. 2006. "Subjective and Objective Measures of Organizational Performance." In *Public Service Performance*, edited by G. A. Boyne, K. J. Meier, L. J. O'Toole, and R. M. Walker, 14–34. Cambridge: Cambridge University Press.
- Barrows, S., M. Henderson, P. E. Peterson, and M. R. West. 2016. "Relative Performance Information and Perceptions of Public Service Quality." *Journal of Public Administration Research and Theory* 26 (3): 571–583. doi:10.1093/jopart/muw028.
- Bifulco, R., and H. F. Ladd. 2005. "Institutional Change and Coproduction of Public Services." *Journal of Public Administration Research and Theory* 16 (4): 553–576. doi:10.1093/jopart/muj001.
- Bohte, J., and K. J. Meier. 2000. "Goal Displacement: Assessing The Motivation For Organizational Cheating." *Public Administration Review* 60 (2): 173–182. doi:10.1111/puar.2000.60.issue-2.

- Bouckaert, G., S. Van de Walle, and J. K. Kampen. 2005. "Potential for Comparative Public Opinion Research in Public Administration." *International Review of Administrative Sciences* 71 (2): 229–240. doi:[10.1177/0020852305053882](https://doi.org/10.1177/0020852305053882).
- Brinkerhoff, D. W., A. Wetterberg, and E. Wibbels. 2018. "Distance, Services, and Citizen Perceptions of the State in Rural Africa." *Governance* 31 (1): 103–124. doi:[10.1111/gove.2018.31.issue-1](https://doi.org/10.1111/gove.2018.31.issue-1).
- Brown, B., and W. N. Benedict. 2002. "Perceptions of the Police: Past Findings, Methodological Issues, Conceptual Issues and Policy Implications." *Policing: an International Journal of Police Strategies & Management* 25 (3): 543–580. doi:[10.1108/13639510210437032](https://doi.org/10.1108/13639510210437032).
- Campbell, D. T., and D. W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56 (2): 81–105. doi:[10.1037/h0046016](https://doi.org/10.1037/h0046016).
- Carpenter, D. P. 2001. *The Forging of Bureaucratic Autonomy*. Princeton, NJ: Princeton University Press.
- Charbonneau, É., and G. G. Van Ryzin. 2012. "Performance Measures and Parental Satisfaction with New York City Schools." *The American Review of Public Administration* 42 (1): 54–65. doi:[10.1177/0275074010397333](https://doi.org/10.1177/0275074010397333).
- Cheon, O., M. Song, A. M. Mccrea and K. J. Meier. 2019. "Health Care in America: The Relationship between Subjective and Objective Assessments of Hospitals." *International Public Management Journal*, DOI: [10.1080/10967494.2018.1553808](https://doi.org/10.1080/10967494.2018.1553808)
- Cyert, R. M., and J. G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.
- Djellal, F., and F. Gallouj. 2009. *Measuring and Improving Productivity in Services*. Cheltenham, UK: Edward Elgar Publishing.
- Escobar-Lemmon, M., and A. D. Ross. 2014. "Does Decentralization Improve Perceptions of Accountability?" *American Journal of Political Science* 58 (1): 175–188. doi:[10.1111/ajps.12043](https://doi.org/10.1111/ajps.12043).
- Favero, N., and K. J. Meier. 2013. "Evaluating Urban Public Schools." *Public Administration Review* 73 (3): 401–412. doi:[10.1111/puar.12022](https://doi.org/10.1111/puar.12022).
- Favero, N., K. J. Meier, and L. J. O'Toole Jr. 2016. "Goals, Trust, Participation, and Feedback." *Journal of Public Administration Research and Theory* 26 (2): 327–343. doi:[10.1093/jopart/muu044](https://doi.org/10.1093/jopart/muu044).
- Gibbons, S., and O. Silva. 2011. "School Quality, Child Wellbeing and Parents' Satisfaction." *Economics of Education Review* 30 (2): 312–331. doi:[10.1016/j.econedurev.2010.11.001](https://doi.org/10.1016/j.econedurev.2010.11.001).
- Hammond, J., E. Speed, P. Allen, I. McDermott, A. Coleman, and K. Checkland. 2019. "Autonomy, Accountability, and Ambiguity in Arm's-length Meta-governance." *Public Management Review* 21 (8): 1148–1169. doi:[10.1080/14719037.2018.1544660](https://doi.org/10.1080/14719037.2018.1544660).
- Hill, K. Q., and A. Hinton-Andersson. 1995. "Pathways of Representation." *American Journal of Political Science* 39 (4): 924–935. doi:[10.2307/2111663](https://doi.org/10.2307/2111663).
- Hood, C. 2002. "The Risk Game and the Blame Game." *Government and Opposition* 37 (1): 15–37. doi:[10.1111/1477-7053.00085](https://doi.org/10.1111/1477-7053.00085).
- Jakobsen, M. 2012. "Can Government Initiatives Increase Citizen Coproduction?" *Journal of Public Administration Research and Theory* 23 (1): 27–54. doi:[10.1093/jopart/mus036](https://doi.org/10.1093/jopart/mus036).
- James, O. 2011. "Performance Measures and Democracy." *Journal of Public Administration Research and Theory* 21 (3): 399–418. doi:[10.1093/jopart/muq057](https://doi.org/10.1093/jopart/muq057).
- James, O., and A. Moseley. 2014. "Does Performance Information about Public Services Affect Citizens' Perceptions, Satisfaction, and Voice Behaviour?" *Public Administration* 92 (2): 493–511. doi:[10.1111/padm.12066](https://doi.org/10.1111/padm.12066).
- Kelly, J. M. 2003. "Citizen Satisfaction and Administrative Performance Measures." *Urban Affairs Review* 38 (6): 855–866. doi:[10.1177/1078087403038006005](https://doi.org/10.1177/1078087403038006005).
- Kelman, S., and J. N. Friedman. 2009. "Performance Improvement and Performance Dysfunction." *Journal of Public Administration Research and Theory* 19 (4): 917–946. doi:[10.1093/jopart/mun028](https://doi.org/10.1093/jopart/mun028).
- Kirkhaug, R., and K. H. Mikalsen. 2009. "Coping with Autonomy: Antecedents of Interaction between Bureaucrats and Politicians." *Public Management Review* 11 (2): 137–154. doi:[10.1080/14719030802685735](https://doi.org/10.1080/14719030802685735).
- Knott, J. H., and G. J. Miller. 2006. "Social Welfare, Corruption and Credibility: Public Management's Role in Economic Development." *Public Management Review* 8 (2): 227–252. doi:[10.1080/14719030600587455](https://doi.org/10.1080/14719030600587455).
- Kogan, V. 2017. "Administrative Centralization and Bureaucratic Responsiveness." *Journal of Public Administration Research and Theory* 27 (4): 629–646. doi:[10.1093/jopart/mux022](https://doi.org/10.1093/jopart/mux022).
- Koppell, J. G. S. 2005. "Pathologies of Accountability: Ican and the Challenge of "Multiple Accountabilities Disorder." *Public Administration Review* 65 (1): 94–108. doi:[10.1111/puar.2005.65.issue-1](https://doi.org/10.1111/puar.2005.65.issue-1).
- Krause, T., and S. Van Thiel. 2019. "Perceived Managerial Autonomy in Municipally Owned Corporations." *Public Management Review* 21 (2): 187–211. doi:[10.1080/14719037.2018.1473472](https://doi.org/10.1080/14719037.2018.1473472).

- Li, H., Q. Wang, and C. Zheng. 2017. "Interjurisdictional Competition and Intracity Fiscal Disparity across Chinese Prefectural Cities." *Governance* 30 (3): 365–385. doi:10.1111/gove.2017.30.issue-3.
- Lyons, W. E., D. Lowery, and R. H. DeHoog. 1992. *The Politics of Dissatisfaction: Citizens, Services, and Urban Institutions*. Armonk, New York: ME Sharpe.
- Maas, C. J., and J. J. Hox. 2005. "Sufficient Sample Sizes for Multilevel Modeling." *Methodology* 1 (3): 86–92. doi:10.1027/1614-2241.1.3.86.
- Manna, P. 2006. *School's In: Federalism and the National Education Agenda*. Washington, DC: Georgetown University Press.
- Marshall, M. 2006. "Parent Involvement and Educational Outcomes for Latino Students." *Review of Policy Research* 23 (5): 1053–1076. doi:10.1111/ropr.2006.23.issue-5.
- Marvel, J. D. 2016. "Unconscious Bias in Citizens' Evaluations of Public Sector Performance." *Journal of Public Administration Research and Theory* 26 (1): 143–158.
- Merton, R. K. 1940. "Bureaucratic Structure and Personality." *Social Forces* 18 (4): 560–568. doi:10.2307/2570634.
- Monroe, A. D. 1998. "Public Opinion and Public Policy, 1980–1993." *Public Opinion Quarterly* 62 (1): 6–28. doi:10.1086/297828.
- Morgeson, F. V. 2013. "Expectations, Disconfirmation, and Citizen Satisfaction with the US Federal Government." *Journal of Public Administration Research and Theory* 23 (2): 289–305. doi:10.1093/jopart/mus012.
- Neshkova, M. I. 2014. "Does Agency Autonomy Foster Public Participation?" *Public Administration Review* 74 (1): 64–74. doi:10.1111/puar.12180.
- Nielsen, P. A. 2014. "Performance Management, Managerial Authority, and Public Service Performance." *Journal of Public Administration Research and Theory* 24 (2): 431–458. doi:10.1093/jopart/mut025.
- Nõmm, K., and T. Randma-Liiv. 2012. "Performance Measurement and Performance Information in New Democracies: A Study of the Estonian Central Government." *Public Management Review* 14 (7): 859–879. doi:10.1080/14719037.2012.657835.
- OECD (Organization for Economic Cooperation and Development). 2015. "PISA 2015 Technical Report." Accessed 26 August 2018. <http://www.oecd.org/pisa/data/2015-technical-report/>
- Olsen, A. L. 2015. "Citizen (Dis) Satisfaction: An Experimental Equivalence Framing Study." *Public Administration Review* 75 (3): 469–478. doi:10.1111/puar.12337.
- Olsen, A. L. 2017. "Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations." *Journal of Public Administration Research and Theory* 27 (4): 562–580. doi:10.1093/jopart/mux023.
- Overman, S. 2016. "Great Expectations of Public Service Delegation." *Public Management Review* 18 (8): 1238–1262. doi:10.1080/14719037.2015.1103891.
- Overman, S. 2017. "Autonomous Agencies, Happy Citizens?" *Governance* 30 (2): 211–227. doi:10.1111/gove.12207.
- Parrado, S., G. Van Ryzin, T. Bovaird, and E. Löffler. 2013. "Correlates of Co-Production: Evidence from a Five-Nation Survey of Citizens." *International Public Management Journal* 16 (1): 85–112. doi:10.1080/10967494.2013.796260.
- Petrovsky, N., J. Y. Mok, and F. León-Cázares. 2017. "Citizen Expectations and Satisfaction in a Young Democracy." *Public Administration Review* 77 (3): 395–407. doi:10.1111/puar.12623.
- Rindermann, H. 2007. "The G-factor of International Cognitive Ability Comparisons: The Homogeneity of Results in PISA, TIMSS, PIRLS and IQ-tests across Nations." *European Journal of Personality* 21 (5): 667–706. doi:10.1002/(ISSN)1099-0984.
- Schneider, M., P. Teske, and M. Marshall. 2002. *Choosing Schools*. Princeton, NJ: Princeton University Press.
- Song, M., and K. J. Meier. 2018. "Citizen Satisfaction and the Kaleidoscope of Government Performance." *Journal of Public Administration Research and Theory* 4 (19): 489–505. doi:10.1093/jopart/muy006.
- Stipak, B. 1979. "Citizen Satisfaction with Urban Services." *Public Administration Review* 39 (1): 46–52. doi:10.2307/3110378.
- Tiebout, C. M. 1956. "A Pure Theory of Local Expenditures." *The Journal of Political Economy* 64 (5): 416–424. doi:10.1086/257839.
- Van Ryzin, G. G. 2004. "Expectations, Performance, and Citizen Satisfaction with Urban Services." *Journal of Policy Analysis and Management* 23 (3): 433–448.
- Van Ryzin, G. G. 2015. "Service Quality, Administrative Process, and Citizens' Evaluation of Local Government in the US." *Public Management Review* 17 (3): 425–442. doi:10.1080/14719037.2013.841456.

- Van Ryzin, G. G., S. Immerwahr, and S. Altman. 2008. "Measuring Street Cleanliness." *Public Administration Review* 68 (2): 295–303. doi:[10.1111/j.1540-6210.2007.00863.x](https://doi.org/10.1111/j.1540-6210.2007.00863.x).
- Van Thiel, S. 2001. *Quangos: Trends, Causes and Consequences*. Aldershot, UK: Ashgate Publishing.
- Verhoest, K., S. Van Thiel, G. Bouckaert, and P. Lægreid. 2012. *Government Agencies: Practices and Lessons from 30 Countries*. New York, NY: Palgrave Macmillan.
- Walker, R. M., M. J. Lee, O. James, and S. M. Ho. 2018. "Analyzing the Complexity of Performance Information Use." *Public Administration Review* 78 (6): 852–863. doi:[10.1111/puar.12920](https://doi.org/10.1111/puar.12920).
- Wynen, J., K. Verhoest, E. Ongaro, and S. Van Thiel, in cooperation with the COBRA network. 2014. "Innovation-oriented Culture in the Public Sector: Do Managerial Autonomy and Result Control Lead to Innovation?" *Public Management Review* 16 (1): 45–66. doi:[10.1080/14719037.2013.790273](https://doi.org/10.1080/14719037.2013.790273).

Appendix

Table A1. Factor-analytical results of parents' evaluation items.

Survey items	Factor loading
Most of my child's school teachers seem competent and dedicated.	0.73
Standards of achievement are high in my child's school.	0.65
I am happy with the content taught and the instructional methods used in my child's school.	0.80
I am satisfied with the disciplinary atmosphere in my child's school.	0.74
My child's progress is carefully monitored by the school.	0.80
My child's school provides regular and useful information on my child's progress.	0.73
My child's school does a good job in educating students.	0.82
Eigenvalues	4.00
Cronbach's alpha	0.87
N	62,916

Table A2. Factor-analytical results of autonomy items.

Autonomy item	Standard autonomy	Human Resources autonomy	Financial autonomy
Selecting teachers for hire	0.19	0.70	0.32
Firing teachers	0.15	0.75	0.27
Establishing teachers' starting salaries	0.18	0.84	0.04
Determining teachers' salary increases	0.16	0.84	0.05
Formulating the school budget	0.09	0.29	0.73
Deciding on budget allocations within the school	0.17	0.09	0.83
Establishing student disciplinary policies	0.63	0.08	0.35
Establishing student assessment policies	0.74	0.15	0.18
Approving students for admission to the school	0.50	0.10	0.16
Choosing which textbooks are used	0.68	0.19	0.15
Determining course content	0.75	0.25	0.00
Deciding which courses are offered	0.68	0.29	0.12
Eigenvalues	2.85	2.77	1.62
Cronbach's alpha	0.80	0.84	0.59
N	437,785	437,785	437,785

Table A3. Descriptive statistics for all variables.

Variable	Mean	SD	Min	Max
Parents' evaluations	0.00	1.00	−4.23	1.75
Student performance	0.00	1.00	−3.58	3.26
School performance	0.00	1.00	−3.79	2.12
Standard autonomy	0.20	1.00	−3.26	2.27
Human resources autonomy	−0.30	0.99	−2.04	2.31
Financial autonomy	0.02	0.91	−3.65	2.22
Highest parent education	4.26	1.60	0	6
Immigrant status	0.18	0.50	0	2
Grade repetition (yes = 1; no = 0)	0.16	0.37	0	1
Ownership (public = 1; private = 0)	0.67	0.47	0	1
Government funding (%)	76.37	31.52	0	100
Number of students (logged)	6.64	0.75	1.10	8.88
Full-time teachers (%)	88.50	17.94	0	100
Student-teacher ratio	15.04	8.36	1.00	96.25

N = 62,916. Immigrant status is coded as native = 0; second generation = 1; first generation = 2. Highest parent education is coded as none = 0; primary education = 1; lower secondary education = 2; upper secondary education = 3; Post-secondary non-tertiary education = 4; tertiary-type B education = 5; tertiary type-A education & advanced research programs = 6.