



AARHUS UNIVERSITY



# Coversheet

---

**This is the accepted manuscript (post-print version) of the article.**

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

**How to cite this publication**

Please cite the final published version:

Blom-Hansen, J., Houlberg, K., & Serritzlew, S. (2021). Jurisdiction Size and Local Government Effectiveness: Assessing the Effects of Municipal Amalgamations on Performance. *European Journal of Political Research*, 60(1), 153-174. <https://doi.org/10.1111/1475-6765.12394>

## Publication metadata

**Title:** Jurisdiction Size and Local Government Effectiveness: Assessing the Effects of Municipal Amalgamations on Performance  
**Author(s):** Jens Blom-Hansen, Kurt Houlberg, Søren Serritzlew  
**Journal:** *European Journal of Political Research*, 60(1), 153-174  
**DOI/Link:** <https://doi.org/10.1111/1475-6765.12394>  
**Document version:** Accepted manuscript (post-print)

**General Rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

JURISDICTION SIZE AND LOCAL GOVERNMENT EFFECTIVENESS:  
ASSESSING THE EFFECTS OF MUNICIPAL AMALGAMATIONS ON  
PERFORMANCE

By

Jens Blom-Hansen, Department of Political Science, University of Aarhus ([jbh@ps.au.dk](mailto:jbh@ps.au.dk))

Kurt Houlberg, VIVE – The Danish Center for Social Science Research ([kuho@vive.dk](mailto:kuho@vive.dk))

Søren Serritzlew, Department of Political Science, University of Aarhus ([soren@ps.au.dk](mailto:soren@ps.au.dk))

Keywords: Municipal amalgamations; mergers; effectiveness; scale effects; optimal  
jurisdiction size

*Abstract.* The optimal jurisdiction size has been debated since Plato and Aristotle. A large literature has studied economic and democratic scale effects, but we have almost no knowledge of the effects of jurisdiction size on the effectiveness of local services. This is due to two methodological problems. First, selection bias and reverse causality often render change in jurisdiction size an endogenous variable. Second, there is a lack of empirical indicators of effectiveness, and most studies therefore focus on spending measures. Extant research thus studies *economies* of scale, leaving *effectiveness* of scale unexamined. We address both problems in a quasi-experimental study of public schools. Our findings from the school area indicate that jurisdiction size does not have systematic effects on effectiveness. Our analysis therefore supports recent studies of economic and democratic scale effects that indicate that the search for the optimal jurisdiction size is futile.

## INTRODUCTION<sup>1</sup>

Over the past half century, local government amalgamation reforms have swept the developed world. Across Europe, North America, and the Antipodes, small jurisdictions have been merged to create more consolidated systems. These reforms have been carried out with various objectives, including strengthening local democracy, making local service provision more efficient, and improving local service effectiveness (Baldersheim & Rose 2010; Fox & Gurley 2006; Kersting & Vetter 2003).

A large body of literature has investigated the democratic potential of local government amalgamations (Dahl & Tufte 1973; Denters et al. 2014; Lassen & Serritzlew 2011; Warren 2011). An equally large literature has analyzed the economic potential—scale effects—of increasing local jurisdiction size (Allers & Geertsema 2016; Blom-Hansen et al. 2014; Blom-Hansen et al. 2016; Byrnes & Dollery 2002; Cobban 2017; Hirsch 1959; Oates 1972; Tiebout 1956; Treisman 2007).

Far less scholarly attention has been devoted to the alleged potential of amalgamations to improve the effectiveness of local services (or effectiveness of scale), despite the near consensus that effectiveness, quality, and service standards are vital concepts to consider when evaluating public sector performance (Boyne et al. 2003: 17–18; Walsh 1991). These aspects figure prominently in political debates: Reformers readily promise local service improvements (Baldersheim & Rose 2010: 12; Kersting & Vetter 2003: 346). They are also discernable in the theoretical debate on optimal jurisdiction size. But they have not been

---

<sup>1</sup> The authors would like to thank Poul Erik Mouritzen, Larry Rose, and three anonymous referees for valuable comments on various drafts. In addition we thank Søren Kjær Foged and Karl Fritjof Krassel for sharing data and Emily Tangsgaard Christensen and Mads Ulrich Matthiessen for vital help preparing and analyzing individual-level register data.

systematically evaluated empirically. Empirical studies of the economic effects of local government amalgamations focus on the effects of mergers on expenditures per capita and are thus limited to *economies* of scale. The important question of *effectiveness* of scale is left unexamined (i.e., the effect of jurisdiction size on the ability of local services to fulfill political goals). In fact, we have no solid knowledge of the effects of local government mergers on local service effectiveness. This is probably due to this concept being notoriously difficult to define, both theoretically and empirically.

The purpose of this paper is to take effectiveness and local government amalgamations seriously by asking: Does increased jurisdiction size enhance public service effectiveness? We start with a conceptual discussion of the effectiveness concept and then evaluate the theoretical discussion of optimal jurisdiction size. We show that effectiveness is and ought to be included in this debate. In the end, our argument is that the effects of jurisdiction size on effectiveness are theoretically inconclusive and that amalgamations therefore cannot be expected to systematically influence effectiveness.

We then turn to testing the argument, which faces three methodological challenges. The first is endogeneity caused by the facts that amalgamations are rarely distributed randomly across local governments. We argue that the Danish 2007 municipal amalgamation reform constitutes a quasi-experiment that effectively addresses this concern. The second challenge is that a relevant case is required to make a convincing test of our zero-effect argument (Gerring 2007: 115–122). We argue that Danish public schools constitute such a case. The third challenge is the indicator problem (i.e., finding a valid measure of effectiveness). Measuring the overall effectiveness of multi-purpose entities is probably a futile task. However, focusing on selected service areas with valid effect measures may be indicative of the overall

effectiveness. Following Andrews et al. (2006), we argue that school exam results are such a measure because they focus on outcomes (not inputs), are measured independently of the organization, are verified, and cover the organization's major goal.

In Denmark, public schools are governed by the local authority: the municipalities. Hence, although our study is limited to the school area, we can simultaneously solve the challenges of endogeneity, relevant case selection, and the indicator problem. Our findings confirm that jurisdiction size does not have systematic effects on effectiveness. This result is in line with recent analyses of democratic and economic scale effects that also fail to identify strong effects of increasing jurisdiction size (Blom-Hansen et al. 2014; Blom-Hansen et al. 2016; Byrnes & Dollery 2002; Denters et al. 2014; Lassen & Serritzlew 2011). In the conclusion, we discuss the implications for the debate on optimal jurisdiction size.

## EFFECTIVENESS IN LOCAL SERVICE PROVISION

Terms like public service quality, performance, improvement, and effectiveness are often used interchangeably. These concepts are vague and contested due to their multidimensional nature and the number of stakeholders taking an interest in public services. Though performance is often linked to the outputs and outcomes of public service provision, it sometimes refers to elements in the entire chain from input to outcome (Van Dooren et al. 2015). Within the last 10–15 years, however, increasing scholarly attention has been paid to conceptual clarification and there is a growing consensus that effectiveness should be understood as the meeting of organizational goals (Andersen et al. 2016; Boyne 2002, 2003a,

2003b; Boyne & Walker 2005; Boyne et al. 2006; Walker & Andrews 2015; Walker et al. 2010; Walker et al. 2018).

Numerous models have been developed that have contributed to clarifying the effectiveness concept. One is the 3Es model, which focuses on the economy, efficiency, and effectiveness of public services. Economy is the cost of producing public services; efficiency is technical (unit costs) or allocative (responsiveness to public preferences); and effectiveness refers to the actual achievement of formal service objectives (Boyne 2002; Walker & Andrews 2015; Walker et al. 2010). Drawing inspiration from this model, we henceforth use the term effectiveness when dealing with the service-improving aspect of amalgamation reforms. We thus preserve the term ‘scale effectiveness’ for the effect of jurisdiction size on the *outcomes* of local government service provision, whereas ‘scale efficiency’ relates to *outputs* and ‘scale economy’ to *inputs* in terms of expenditures per capita and means of production. Along this conceptual line, improved outcomes at constant inputs would imply an increased cost-effectiveness (Van Dooren et al. 2015).

The 3Es model separates effectiveness from costs, unit costs, and responsiveness and places emphasis on the effects of service provision in relation to service objectives. This is an important first step in conceptual clarification. But the term ‘service objectives’ is still rather abstract. An important distinction is between objective and subjective measures of service objectives (Andersen et al. 2016; Andrews et al. 2006; Walker et al. 2018). Subjective measures can be internal (e.g., performance evaluations by employees or managers) or external (e.g., judgements by consumers or outside inspectors). They are advantageous in that they respect that service goals can be manifold and that various stakeholders place different emphasis on different goals (Pollitt 2013). However, subjective measures suffer from the

weakness that they may reduce effectiveness to the subjective impressions of powerful stakeholders. Full reliance on subjective measures can thus easily lead to relativism.

Accordingly, objective measures are often seen as superior to subjective measures (Andersen et al. 2016; Meier & O'Toole 2013; Walker et al. 2018). However, some service objectives are inherently subjective, e.g. parent satisfaction in kindergartens or public trust in the police. The choice between subjective and objective indicators should therefore depend on the service objective in question.

No matter whether objective or subjective measures are used, effectiveness is measured against one or several service objectives. This understanding of effectiveness corresponds to what Boyne (2003a) calls the goal model, which assumes that formal organizations seek to reach objectives that citizens cannot reach on their own. This assumption draws on classical thoughts on organizational goals developed in organization theory (Cyert & March 1992 [1963]: 30–52; Mohr 1973). It does not imply that organizations are unitary actors and that goal conflicts are absent, but rather that goals are constraints imposed on the organization via a bargaining process among coalitions within the organization. Goals may be adapted over time to changes in the coalition structure, but in the absence of dramatic shifts, adaptations are gradual. In the shorter run, it is possible to treat goals as given. In this understanding, increased effectiveness means that organizational goals are met to a higher degree.

As noted by Boyne (2003a: 222), the strength of focusing on goals is that they emphasize tangible elements in the content of services (e.g., speed, quality, reliability) that are likely to be valued by all stakeholders even if they may disagree about the weights that should be attached to them. The extent to which these tangible elements are met constitutes the organization's effectiveness.

To measure effectiveness in practice, indicators are needed that meet a number of criteria (Andrews et al. 2006; Boyne 2003b; Walker & Andrews 2015). First, indicators must focus on the products delivered by the organization; that is, indicators must be on the outcome side of the organization. Second, indicators must be impartial, independent, and detached from the organization; that is, it should not be possible for an organization to influence the indicators in other ways than by changing its effectiveness. Third, indicators should be open to scrutiny and include an external process to verify their accuracy; that is, some external validity check should be involved. Finally, indicators should be comprehensive and cover the organization's major goal; that is, indicators should not focus on a selected sub-aspect of the organization's goals. In that case, stakeholders will disagree on the usefulness of the indicator.

This list of criteria may seem demanding, but in some economically important and politically salient service areas, many measures in fact meet them. Andrews et al. (2006: 16) use the example of school exam results, which in their view constitute a satisfactory measure of school effectiveness because exam results 'reflect an element of the effectiveness of schools, and students' achievements are validated through the grading of their work by external examiners... [and exam results are] publicly available and open to scrutiny by the community.' This specific indicator is frequently used to measure effectiveness, for example by Meier and O'Toole in their influential work on the school system in Texas (e.g., Meier & O'Toole 2006).

In sum, we believe that it is possible to define effectiveness, both theoretically and empirically. We therefore also deem it possible to investigate the relationship between municipal amalgamations and the effectiveness of local public services. We start this



investigation in the next section by discussing the theoretical relationship between jurisdiction size and public service effectiveness.

## JURISDICTION SIZE AND EFFECTIVENESS IN THEORY

The optimal size of political systems has been a recurrent question in political debates since Plato and Aristotle. This debate often addresses the relative merits of democratic and economic concerns (Dahl & Tufte 1973; Hooghe & Marks 2009; Treisman 2007). Democratic concerns like participation, trust, responsiveness, and accountability may be served best in small jurisdictions. However, equally important democratic concerns like diversity in beliefs and values, media coverage, and active community groups may be served best in large jurisdictions (Dahl & Tufte 1973; Lassen & Serritzlew 2011; Warren 2011). Turning to economic concerns, small jurisdictions can better tailor public services to local preferences if preferences vary geographically or if citizens can easily move to localities offering more attractive tax-service packages. However, large jurisdictions may be economically more efficient because of economies of scale (Allers & Geertsema 2016; Blom-Hansen et al. 2014; Blom-Hansen et al. 2016; Hirsch 1959; Oates 1972; Tiebout 1956).

While the debate thus normally focuses on democratic and economic parameters, upon closer inspection it also includes concerns regarding public service effectiveness, although this concern is often couched in economic terms. This may be justified if effectiveness is seen as the ultimate effect in a longer chain relationship. For example, one might argue that large jurisdictions are more efficient because they can harvest scale effects, thereby improving public services without raising taxes. Scale effects in large jurisdictions might therefore be

reflected in better services rather than lower costs. In this example, size impacts efficiency, which in turn impacts effectiveness. However, there are also arguments that size may have a more direct effect on effectiveness. We discuss these arguments in more detail below. We want to demonstrate that effectiveness should be considered a separate concern in the debate on optimal jurisdiction size, although—like democratic and economic concerns—upon closer inspection, it does not unequivocally speak for a certain jurisdiction size.

We begin with Dahl and Tufte's (1973) influential work. They argue that size may affect a political unit's system capacity, whereby they mean the system's capacity to respond to the collective preferences of citizens. While this is not identical to our definition of effectiveness, it is close enough to merit discussion. Dahl and Tufte (1973: 110–117) argue that two mechanisms may link size to capacity. The first is specialization: Greater population facilitates greater specialization. This not only allows for more efficient service production—thereby creating economic scale effects—it also allows the use of more highly skilled labor. Since skilled labor is an important precondition for the production of high-quality labor-intensive services, it follows that size may be positively related to effectiveness. The second mechanism is the ability to recruit staff with the needed skills. Dahl and Tufte (1973: 117) discuss the example of hiring artists for a first-rate opera house, which may be easier in New York City than in other subnational units in the USA. This argument can easily be extended to other lines of production, since many public servants prefer to work in stimulating professional environments, which may be easier to create in large jurisdictions.

Dahl and Tufte's two mechanisms pervade much of the debate on size and performance. In the USA, the progressive reform movement, a consistent advocate of municipal consolidation, has been a strong proponent of the specialization argument. One of this movement's core

arguments is that large jurisdictions make it possible to apply professional expertise to the complexities of modern society (Knott & Miller 1987; Ostrom 1972). In the UK, large-scale municipal amalgamations were undertaken in the 1960s and 1970s because, *inter alia*, large jurisdictions were expected to be able to employ specialist staff and attract higher caliber politicians (Boyne 1996). Similarly, local government amalgamation reforms in a number of European countries in recent decades have been motivated by the need for professional service provision and the capacity to meet the demands of modern citizens for high quality (Baldersheim & Rose 2010).

To Dahl and Tufte's two mechanisms linking size and effectiveness we can add a third: the optimal division of the local jurisdiction into catchment areas. Many local services, such as primary schools, libraries, and residential homes for the elderly, are not delivered by the local government, but by institutions operating in separate service districts under the local government. The difference between the local government and its catchment areas resembles the distinction drawn by economists between firms and plants, and firm-level and plant-level efficiency (Sawyer 1991). Plant-level effectiveness—effectiveness at the level of catchment areas—does not necessarily depend on jurisdiction size, since the size of schools, libraries, residential homes, etc., and the districts of these institutions may be changed within a given jurisdiction. It may nevertheless depend on jurisdiction size, however, as the optimal boundaries of these districts may be easier to draw in large jurisdictions. But firm-level effectiveness—effectiveness at the local-government level—is probably more likely to depend on jurisdiction size, since large jurisdictions are arguably better able to flexibly deploy expert managers and administrative support across catchment areas. Increased jurisdiction size may thus improve the firm-level management of institutions at the plant level even though the sizes of institutions remain unchanged.

In sum, it seems justified to add an effectiveness dimension to the debate about optimal jurisdiction size, since several mechanisms may link jurisdiction size to effectiveness: specialization, recruitment power, management, and within-jurisdiction districting. However, just as democratic and economic concerns can speak both for and against large jurisdictions, effectiveness concerns favor not only large jurisdictions. In fact, several effectiveness arguments can be brought forward that favor small jurisdictions.

First, large jurisdictions will be fewer in number, which reduces competition and leads to less experimentation in the public sector. This is a common argument in the economic literature on jurisdiction size (Oates 1972: 12; Tiebout 1956; Treisman 2007: 222–235). But less experimentation would hurt not only efficiency, as argued by economists, but also effectiveness.

Second, one should keep in mind that the potential effectiveness gains of large jurisdictions are not harvested automatically (cf. Treisman 2007: 58); doing so requires political action. A parallel may here be made to Boyne's (1998: 133–135) theorizing on the effects of contracting out. He argues that the cost-savings from contracting out are not likely to be fully returned to the taxpayers, but retained instead by local government. If senior officials are budget-maximizers and face low political scrutiny, they will spend the savings from contracting out on other services rather than returning them to local citizens in the form of reduced taxation. A similar argument can be made in relation to effectiveness. Harvesting the effectiveness potential from amalgamation may require unpopular adjustments, such as laying-off existing personnel or redistricting catchment areas. In the absence of specific incentives to carry out these unpopular adjustments, the effectiveness potential is not likely to

be harvested; in other words, much depends on the context and specificities of the local government in question.

Third, upon closer inspection, arguments concerning the importance of jurisdiction size are often actually arguments regarding urbanity. This also holds for certain effectiveness arguments. Take Dahl and Tufte's example of hiring artists for the New York Opera. Is this facilitated by the size of the New York City government or the fact that New York is a massive urban center? Disentangling the importance of size and urbanity is of paramount importance for reformers. By amalgamating a number of small rural municipalities, reformers do not create a large urban municipality, but a large rural one. If effects depend on urbanity, no effects will materialize from such reform.

Finally, the effectiveness of local services is a multi-causal phenomenon. Boyne (2003b) and Walker and Andrews (2015) provide overviews of studies of the determinants of local service effectiveness such as resources, regulation, markets, management, organization, strategy, planning, and networking. In both overviews, the main determinants are found to be factors like resources, management, and planning—not organization. This does not necessarily mean that jurisdiction size is unimportant, but it does mean that it is unlikely to have a systematic, consistent effect. Again, much seems to depend on context.

We conclude that the effectiveness arguments favoring large jurisdictions are only likely to hold in certain contexts. In other contexts, small jurisdictions may be more favorable to effectiveness. In most circumstances, however, factors other than jurisdiction size are likely to matter more for effectiveness. This may be disappointing news to reformers, as the most sober conclusion to draw is that, upon closer inspection, the link between jurisdiction size and

effectiveness is weak and unsystematic. Amalgamating local jurisdictions is unlikely to have any systematic impact on local public service effectiveness.

## DESIGN: EFFECTIVENESS OF SCALE IN THE MUNICIPAL SCHOOL SYSTEM

When testing our argument, we face three methodological challenges. The first is endogeneity. We argue that the Danish 2007 municipal amalgamation reform constitutes a quasi-experiment that addresses this challenge. Second, in order to make a sound test of our zero-effect expectation, we need a relevant case, that is, a case where it is likely that politicians will make an effort to harvest effectiveness gains. Since this may require unpopular organizational adjustments, an electoral incentive is needed to make effectiveness gains a likely outcome. This again requires that voters care about the service in question. We argue that Danish public schools constitute such a case, as public education is one of the most economically important and politically salient services. Although there are limits to the generalizability of the findings from such a case, a zero-finding would constitute relatively strong disconfirmation of arguments on scale effects, at least for comparable welfare areas where service is delivered by institutions at the sub-municipal level, e.g. kindergartens, libraries, and institutions for the elderly and handicapped. Third, we need to solve the indicator problem. We argue that school exam results can do this. Since Danish public schools are governed by the municipalities, we are in a unique position to address all three methodological challenges simultaneously. We discuss the three challenges in more detail below.

First, the endogeneity problem has two sources: self-selection and simultaneity (reverse causality). Self-selection creates endogeneity bias when municipalities can decide for themselves whether to merge. These decisions may be affected by a host of factors that may also affect local government effectiveness. An example: Municipalities in economic trouble may be more likely to merge with other municipalities. They may also produce output less effectively because economic troubles led them to lower public service standards. This leads to a spurious negative correlation between mergers and effectiveness, even if size has no causal effect on effectiveness. Simultaneity occurs when the dependent variable affects the independent variable. This is a concern because effectiveness may affect the likelihood of municipal mergers. One mechanism could be that highly efficient municipalities are more attractive as amalgamation partners. In this case, a correlation between size and effectiveness might possibly reflect an effect of effectiveness on size; not the causal effect of size on effectiveness.

We use a quasi-experiment to address this problem. The 2007 Danish municipal amalgamation reform is a source of exogenous variation in jurisdiction size.<sup>2</sup> Before 2007, Denmark had 271 municipalities, 239 of which were merged to form 66 municipalities (the treatment group), while 32 municipalities (the control group) remained unaffected by the reform. This reform, which has been used to study other effects of jurisdiction size (Lassen & Serritzlew 2011; Blom-Hansen et al. 2014; Foged 2016; Blom-Hansen et al. 2016), was decided and implemented by the central government. With a few exceptions for islands, all smaller local governments were required to amalgamate with neighboring municipalities. The assignment to treatment is obviously not random, but the central question to the design is whether the dependent variable is likely to have affected individual local governments'

---

<sup>2</sup> This description of the Danish municipal reform builds upon Blom-Hansen et al. (2016).

merger choices. Studies have shown that factors such as local identity (Mouritzen 2006) and social connections between geographic areas dominated these decisions (Bhatti & Hansen 2011). We conclude, in line with the cited studies, that the decisions to amalgamate were exogenous to the economic and organizational features of the municipalities.

Second, we identify public schools in Denmark as a relevant case for scale effectiveness of municipal amalgamations. Within the discipline of education economics, many studies of the relationship between school district size and school quality have been made. But findings are mixed. Review studies conclude that studies find both positive, negative and no effects of district size on school quality (Fox 1981; Andrews et al. 2002). More recent studies seem to confirm this state of affairs. Heinesen (2005) and Sandsør et al. (2017) find positive effects of size, while Gordon and Knight (2008) and Berry and West (2010) find no or weak size effects. So, in general, it is difficult to say whether the school area is a more-likely or less-likely case for scale effectiveness. Much seems to depend on local context.

We contend that public schools in Denmark constitute a relevant and likely case. This is due to the unusual political salience of public schools in Denmark. Survey studies document that of all the functions of Danish multi-purpose municipalities, the school area is consistently highly politically salient and constitutes the top-ranked service in the eyes of both local politicians and citizens (Bhatti et al. 2017; Houlberg et al. 2016). Hence, if amalgamations give rise to scale effects, the improved performance of schools is more likely to be politically prioritized in Denmark than improvements of comparable services, e.g. kindergartens, libraries, and institutions for the elderly and handicapped.<sup>3</sup> Although it would obviously be

---

<sup>3</sup> Consistent with this, Heinesen (2005) finds a positive relationship between population size and education attainment in Danish public schools. However, as this study was conducted before the amalgamation reform of 2007 it is based on non-experimental data.



desirable to investigate all these areas, a zero-finding in the school area in Denmark would therefore constitute relatively strong disconfirmation of scale effects.

Third, as argued above, solving the indicator problem requires meeting four criteria. Andrews et al. (2006: 16) use school exam results as an example of a good measure of effectiveness (see also Andersen et al. 2014, 2016). We agree and contend that exam results in Danish public schools satisfy each of the four criteria. First, a good indicator should focus on outcomes, not inputs. School exam results are measures of the outcomes of schools. Input factors, such as expenditures, number of teachers, teaching materials, etc. are not part of the measure. Second, the indicator should be impartial and measured independently. School exams in Danish public schools are formal and based on criteria set in national curricula. This ensures that individual schools can only hope to affect the indicator by changing school effectiveness. Third, the indicator should be scrutinized by external processes. Exam results are decided by the local examiner and an independent external examiner appointed by the Ministry of Education. Fourth, the indicator should cover the organizations' major goal. As explained below, academic achievement is the most important objective of the Danish public school system, and exam results are designed to measure this.

However, we do not argue that exam results measure all aspects of performance in Danish schools. A critique that is often raised against exam results is that they do not measure how effective schools are at developing character, integrity, and empathy for others or at teaching students citizenship skills (Koretz 2008; Ravitch 2013). This critique is also relevant in the Danish case, and a complete measure of school effectiveness would include these aspects. However, this is not possible due to lack of data. But we think that proceeding with an analysis focusing on exam results is justified on the ground that improvement of academic

achievement is a central goal for Danish schools and has been a core aim in Danish national education policies for decades (cf. below).

In sum, the Danish municipal amalgamation reform of 2007 combined with data on public school performance allow us to effectively address the methodological challenges facing a study of the causal effects of jurisdiction size on effectiveness.

## PUBLIC PRIMARY SCHOOLS IN DANISH LOCAL GOVERNMENT

Public primary schools in Denmark offer primary and lower secondary education consisting of a one-year pre-school class, grades 1–9 and an optional 10<sup>th</sup> grade. Parents have the right to enroll their children in the public school in their local school district, but they can also—subject to availability—opt for another public school or pay for enrolment in a private school. About 80% of all children are enrolled in public schools (Andersen et al. 2014).<sup>4</sup>

All 98 municipalities operate public primary schools, but are also responsible for local regulations (e.g., environmental regulations according to national law), infrastructure (roads, garbage collection, utilities), and local welfare services. The municipalities are governed by elected city councils and mayors elected from among the councilors.

Within the framework set by the national Education Act, the municipalities are granted full organizational and budgetary responsibility for public primary schools. The municipalities decide the number and size of the schools. As of 2017, the average size of the 1,091 public

---

<sup>4</sup> Source: Local government database, Ministry of the Interior.

primary schools is 485 students.<sup>3</sup> Funding is determined by the city council. The average annual cost per student in primary education is about USD 11,355. This is among the highest in the OECD (OECD 2016: 183).

The municipalities are responsible for ensuring that the quality of education meets national objectives. Although schools have other objectives, academic achievement is among the most important goals and its improvement has been a core aim in national education policies since the turn of the century (Andersen et al. 2014). The grades in final exams in seven different subjects at the end of grade 9 are a core indicator of student performance. Some examinations are written and some are oral, but all of them involve at least one external examiner.

In addition to academic achievement, a central goal in Danish education policies since 2006 has been that 95% of each cohort achieves at least an upper secondary or vocational education. Primary schools are therefore also evaluated against the degree to which they provide their students with the competences and motivation to begin and complete an upper secondary education. At the national level, this goal is evaluated by register data on students enrolled in and completing upper secondary schools. We use this as an alternative measure of effectiveness.

## METHODS AND DATA

We seek to estimate the effect of jurisdiction size on the effectiveness of public schools. We use the Danish municipal reform in 2007 as a source of exogenous change in jurisdiction size. Since we can assume that individual merger decisions are unrelated to the performance of

local schools, we can estimate the effect of jurisdiction size by comparing the pre-reform and post-reform performance of public schools in the treatment group of the 66 merged municipalities with the control group of the 32 non-merged municipalities. We observe public school performance in the period 2003–2016 (i.e., four pre-reform years and ten post-reform years). To compare municipalities before and after the reform, we impose the post-2007 structure on the pre-2007 structure by aggregating pre-reform municipalities to their post-reform size. We exclude five islands and three non-merged municipalities with pre-reform, two-tier status as both county and municipality. This gives a total of 1,260 observations—364 observations in the control group, 896 in the treatment group.

We use a difference-in-difference (DiD) approach to identify the causal effect of jurisdiction size (cf. Wooldridge 2009), which allows us to decompose the change in effectiveness for public schools into its components: the change in effectiveness due to increased jurisdiction size and the change in effectiveness due to other factors, such as a general time trend. We do this by estimating this equation:

$$(1) \quad Y_i = \alpha + \beta_1 TG_i + \beta_2 T_i + \beta_3 TG_i \times T_i + \varepsilon_i,$$

where  $Y_i$  denotes effectiveness of public schools in municipality  $i$ ,  $TG_i$  is 1 when municipality  $i$  belongs to the treatment group and 0 if not,  $T_i$  is a time variable (1: post reform; 0 pre reform), and  $TG_i \times T_i$  is an interaction term (which will be 1 for treated municipalities post reform).  $\beta_3$  is the estimate of the effect of jurisdiction size on effectiveness (the estimate is based on treated municipalities, so  $\beta_3$  is actually an estimate of the average treatment effect for treated, ATT).  $\beta_1$  estimates the pre-reform difference between municipalities in the treatment and control groups, and  $\beta_2$  estimates the time trend (post reform vs pre reform) in

effectiveness of public schools. It is straightforward to expand (1) with additional time variables  $T_{i1} - T_{im}$  and corresponding interaction terms to estimate the changes from year  $t = 0$  until  $t = m$ .

In equation (1), treatment in line with other recent studies is measured by a dummy for amalgamation (e.g. Lassen & Serritzlew 2011; Blom-Hansen et al. 2016). However, as economies of scale may be curvilinear (e.g. Jugl 2019; Treisman 2007), we as a robustness test also model treatment by an indicator measuring the linear increase in population size (see Foged 2016) and squared size increase.

As explained above, we use two performance indicators. The primary indicator is exam grades. We use the average grade achieved by each student in seven subjects at the end of grade 9. Grades in each subject are given on a 7-point scale, and the average grades for all students completing at least four of the exams are aggregated to the municipal level. In order to control for potential time and place variation in individual socioeconomic characteristics, we use the average exam grades as well as including an indicator of average exam grades controlled for the individual socioeconomic status (SES) of each student based on their register data (e.g., gender, ethnicity, number of siblings) as well as the parents (e.g., education, income, family status, labor market status). Across the 14 years of analysis, this estimation is based on 540,000 observations (i.e., a yearly average of individual-level data of around 38,500 students and their parents). The second indicator of performance is the share of young people aged 18–21 years either attending or having completed upper secondary education. Both indicators are based on individual-level register data from Statistics Denmark, which we aggregate to the municipal level.

Our core objective is to establish whether increased jurisdiction size leads to higher effectiveness. We are less interested in the potential mechanisms linking jurisdiction size and effectiveness and in potentially moderating organizational and management factors. To shed some light on potential scale dynamics, however, we supplement the two indicators of effectiveness with an indicator of *economies* of scale (measured as cost/student) as well as an indicator of *efficiency* of scale (measured as the number of teaching lessons).

Municipal school costs and performance are influenced by factors other than jurisdiction size. And since the assignment of municipalities to treatment and control groups is not randomized, we include a number of control variables known from previous analyses of primary schools in Danish municipalities (Andersen & Serritzlew 2007; Heinesen 2004). First, we include four indicators of school expenditure needs: dispersed settlements plus three indicators of social problems (social housing, single-parent children, ethnic minority students). Second, an indicator of fiscal pressure controls for variations in economic potential among the municipalities. Third, we control for changes in the number of students, since a fluctuating number of students may make it harder to adjust costs, personnel, and educational activities to the changed fiscal conditions. Fourth, to control for potential differences in political priorities we include a variable measuring the share of socialist seats in the local council. Please see Online Appendix Table A1 for descriptive statistics on all variables.

## EMPIRICAL ANALYSIS

Before turning to the DiD-based regression analyses, we present a first view of the data in Figure 1, which shows the development over time in the average grades for the final exams

and the share of students attending upper secondary education for amalgamated and non-amalgamated municipalities.

First, Figure 1 shows parallel trends for amalgamated and non-amalgamated municipalities before the reform in 2007. This is crucial for the DiD-analyses presented below. Second, there is clearly a time trend that must be accounted for. Figure 1a shows a general increase in grade levels since 2006, both for amalgamated and non-amalgamated municipalities. This is most likely a reflection of grade inflation, which occurred following a reform of the grade scale in 2006 (EVA 2019). Figure 1b shows an general increase in the share of young people in upper secondary education. This trend is most likely related to the fiscal crisis, which hit Denmark around 2008. During economic upturns such as the one in 2004–2007, the uptake in your education tends to be lower. The DiD-based analyses account for these time trends. Third, if the amalgamations affected performance, different trends for amalgamated and non-amalgamated municipalities should occur after the reform. However, the graphs paint no such picture.

[Figure 1]

Based on the graphs, it appears as though the municipal amalgamations had no effect on public school performance. However, this conclusion is preliminary. We need to check that we get the same results when potential confounders are taken into consideration. We therefore now turn to the results of the DiD analyses.

Table 1 compares the average pre- and post-reform levels in the amalgamated and non-amalgamated municipalities in each of the four pre-reform years and ten post-reform years. The single-year estimates make it possible to identify the exact timing of a reform effect. This is important, since reform effects are unlikely to materialize instantaneously, as student performance is not a product of just a single year of schooling.

[Table 1]

Before turning to the indicators of scale effectiveness, we will briefly comment on the results regarding scale economy. The dummy variable *amalgamated* indicates whether the municipality belongs to the treatment group of amalgamated municipalities. It is the estimate of the pre-reform (2003) difference between the treatment and control groups. The estimate indicates no significant difference in costs per student between the two groups in 2003. More importantly, the DiD estimators for the pre-reform years ( $\text{amalgamated} \times 2004$ ,  $\text{amalgamated} \times 2005$ , and  $\text{amalgamated} \times 2006$ ) are insignificant, meaning that the difference in difference between the treatment and control groups did not change significantly from 2003 to 2006 (i.e., the two groups exhibit parallel trends before the reform).

Turning to the DiD estimators for the post-reform years ( $\text{amalgamated} \times 2007$  to  $\text{amalgamated} \times 2016$ ), the estimates are not significantly different from zero in any year. No reform effects or indications of economies of scale for public schools can thus be identified. Model 2 shows that the same result holds for the number of teaching lessons.



However, the key question for our purposes is whether size affects effectiveness or, in other words, whether the potential scale effects did not materialize in the form of lower costs or increased teaching output, but instead as higher effectiveness. The indicators of scale effectiveness in the right side of Table 1 show no sign of scale effects. The estimates for *amalgamated* show that no pre-reform differences existed between the treatment and control groups. In addition, the DiD-estimators for the pre-reform years are generally insignificant and thus signal parallel pre-reform trends in effectiveness. The only exception is the significant 2006-estimate for upper secondary education. This negative estimate indicates that effectiveness decreased in the amalgamated municipalities in the last year before the reform. This may be a reflection of organizational turmoil and focus on preparing the amalgamation and a less-than-usual focus on motivating graduating students to pursue upper secondary education. This interpretation is supported by the fact that the amalgamated municipalities in the first two years after the reform—the constitutive years for the new organization—are also less effective regarding pursuit of upper secondary education. Then, from 2009 onwards, the development in the amalgamated municipalities does not differ from the non-amalgamated municipalities; no signs of long-term scale effectiveness can be identified—neither positive nor negative.

Returning to our main measure of effectiveness, final exam grades, neither of the two indicators shows any systematic sign of scale effectiveness. Regardless of whether we consider the actual grades achieved or grades controlled for individual SES, the DiD estimators show no permanent scale effects. In the first five years after the reform, negative scale effects are indicated in model 3. However, these seem to be of a temporary nature and cease to be significant when controlling for individual-level SES in model 4.<sup>5</sup> The overall

---

<sup>5</sup> Grades controlled for SES are not yet available for 2016.

conclusion of Table 1 is therefore that jurisdiction size does not have any systematic effect on local government effectiveness.

It is worth commenting briefly on the remaining findings in Table 1, the year dummies estimate the general time trend for each year relative to the initial year, 2003. Many of these dummies are statistically significant, not least in the post-reform years. Overall, the time dummies confirm the visual impression from Figure 1, showing a common trend for waning expenditures from 2010, increasing grades from 2007, and an increased share of students attending upper secondary education from 2009. Turning to the control variables, the socio-economic variables have the expected effects and an increased share of socialist seats in the council is reflected in higher per-student spending priorities and lower average student performance.

To check the robustness of the results, we have run models testing for both linear and curvilinear scale effects (see Online Appendix Tables A2 and A3). These robustness tests provide substantially the same results as in table 1, i.e. no sign of scale effects.

In sum, we conclude that increasing the size of municipalities seems to generate neither *economies of scale*, *efficiency of scale*, nor *effectiveness of scale* in the provision of local school services. The fact that the amalgamations did not reduce per capita costs (Blom-Hansen et al. 2016) does thus not seem to be caused by scale effects being transformed to increased effectiveness.

## DISCUSSION AND CONCLUSION

Across the developed world, municipalities have been merged to form larger jurisdictions (Blom-Hansen et al., 2016). The goals of these merger reforms are typically grand (Baldersheim & Rose 2010: 12; Kersting & Vetter 2003: 346). Reformers hope to strengthen democracy, increase service provision efficiency, and improve local service effectiveness.

A large literature has studied how increased jurisdiction size affects democracy and economic efficiency. The classic idea is that there is a trade-off between democratic and economic effects (Dahl & Tufte 1973; Blom-Hansen et al. 2014). In this view, a small jurisdiction is good for democracy (because of the closeness of politicians and citizens) but bad for economic performance (due to economies of scale). The empirical literature suggests that the democratic effects of increased size are negative, albeit small, and that the economic effects are null or very limited. This gives reason for pause for prospective reformers.

However, according to the 3Es model, it is relevant to distinguish between the economy, efficiency, and effectiveness of public services. The first E, the pure economic effects, have to do with the costs of producing services, and the second, efficiency, with unit costs and allocative efficiency. Extant studies focus on these effects and ignore the third E, effectiveness, which refers to whether goals are actually achieved. This is important because the grand goals of merger reforms might be met after all if they lead to better goal attainment. Despite the many studies of service costs and efficiency and near consensus that effectiveness, quality, and service improvement are essential to consider when evaluating performance, we have strikingly little evidence regarding the effectiveness of scale.

The aim of this paper has been to study the causal effect of local government mergers on effectiveness. By doing this, we can get one step closer to clarifying whether there is a trade-off between democratic and economic concerns. In a quasi-experimental study of Danish public schools, we find that the increasing jurisdiction size of local government has no effect on effectiveness. The analyses are based on a DiD-approach to pre- and post-treatment observations for student performance in amalgamated and non-amalgamated Danish municipalities. We combine this dataset with register data on school expenditures, the number of students, exam grades, number of teaching lessons, and socioeconomic characteristics.

By focusing on Danish public schools, we have identified a relevant case; that is, a case where it is likely that politicians will make an effort to harvest effectiveness gains. In addition, we can solve two methodological problems. First, endogeneity is a serious concern since mergers are rarely distributed randomly across local governments (selection bias) and since jurisdiction size may have an impact on performance, but performance is also likely to have an impact on the decision whether to merge (reverse causality). We address the endogeneity problem by using the Danish local government reform in 2007 as a quasi-experiment. The reform constituted an exogenous shock to 239 municipalities, while 32 municipalities were left untouched. Second, there is a lack of empirical indicators of service standards. This implies that most empirical studies focus on spending per capita as the indicator of scale effects. We handle the indicator problem by focusing on public schools, which in Denmark are run by local government. It would obviously be ideal to study several policy areas, but in this study we have prioritized the relevance of the case and the indicator problem. Focusing on schools allows us to study school exams, a widely recognized indicator of effectiveness. We follow the effects for ten years after the reform. This is important because potential effects on student performance are not likely to materialize instantaneously after merging municipalities.

Based on this design, we can be confident that the massive increases in jurisdiction size produced by the Danish amalgamation reform had no causal effect on effectiveness.

We now turn to the discussion of generalizability and broader implications for public policy and offer some reflections on the theoretical implications. Studying public schools in Denmark allows us to address the main methodological problems that have led extant research to focus on economies of scale rather than effectiveness of scale. However, the focus on a single sector of Danish local government also means that one should be cautious when drawing wider conclusions.

First, there is the question of the empirical scope of the findings. Danish local government consists of municipalities within the range of 5,000 to 100,000. It is obviously unfair to claim that increased jurisdiction size can never improve effectiveness. It is likely that jurisdictions can become too small, and based on this study we cannot rule out that economies of scale materialize for extremely large jurisdictions. However, for population sizes within the empirical range of 5,000 to 100,000, the Danish case provides causal evidence for absence of scale effectiveness of running the school system in a multi-purpose jurisdiction.

Second, the argument rests on a political logic: Gains in effectiveness will only be reaped if politicians are rewarded for it. Hence, the political logic depends on the saliency of the service area. The ‘politics matter’ literature has shown that political ideology is an important explanation of public policy (Imbeau et al. 2001; Meier & O’Toole 2006). However, the political preferences have a stronger effect in salient policy areas (Baekgaard et al. 2014). We have argued that the Danish public school sector is highly salient, and that this implies that politicians have a strong incentive to increase effectiveness if at all possible. From this we

would infer that it is, *ceretis paribus*, unlikely that jurisdiction size is associated with effectiveness in less salient policy areas. On the other hand, of course, we cannot from this study draw conclusions for policy areas that are even more salient and important to voters than public schools in Denmark. While we do not have evidence on effectiveness in other service areas, studies exist of the economic costs and efficiency. In Danish local government, the effect of jurisdiction size on costs is null for all major service areas (public schools, daycare, elder care and care for children with special needs). Jurisdiction size have limited effects for costs of administration, roads and the labour market area, and these effects cancel out. We cannot transfer these results to effectiveness, but they do suggest that the school area does not stand out as special when it comes to other economic effects.

Third, the potential for effectiveness also depends on the production function. The school sector is quite complex and involves both plan- and firm-level management. It may well be that it is possible to increase effectiveness in areas with simpler production functions, e.g. city hall administration or local road construction. However, these areas may be less politically salient, and harvesting effectiveness gains may therefore be less politically attractive.

To sum up the policy implications, public school policy is highly salient in Danish local government, and improvement of public schools is popular among both politicians and parents. In other words, this is a policy area that voters care about deeply, and politicians are likely to be rewarded by voters when delivering higher effectiveness. If it is possible to improve performance when jurisdiction size is increased, it should happen here, and we suggest that effectiveness of scale is even less likely in comparable spending areas that are less salient and popular. The policy implication of this is that prospective reformers should

not assume that larger jurisdiction size will improve service provision, or, more bluntly, that the quest for an optimal jurisdiction size is likely to be futile.

To sum up the theoretical implications, the relationship between jurisdiction size and economy is quite complex. First, the 3Es suggest that it is insufficient to study costs and efficiency. Jurisdiction size also has potential effects for effectiveness. Second, gains in effectiveness do not occur automatically, they require political action. This means that the saliency of the policy area is central. Third, the production function is important. Both saliency and production functions vary across service areas. This may mean that it is meaningless to talk about *the* optimal jurisdiction size in local government. If jurisdiction size matter for service production, it does so in different ways for different service areas – for political as well as technical reasons.

The absence of a causal link between jurisdiction size and the effectiveness of public services provides an important contribution to the debate on the optimal size of local jurisdictions. Several studies suggest that democratic scale effects are negative, but small (Lassen & Serritzlew 2011; Denters et al. 2014), and that economic scale effects are, at best, very limited (Blom-Hansen et al. 2014; Blom-Hansen et al. 2016; Byrnes & Dollery 2002). In contrast, the link between jurisdiction size and effectiveness, which figures prominently in practical reform debates, is uncharted research territory. Our study suggests that the results from studies of democratic and economic scale effects also hold for effectiveness: no systematic scale effects can be identified. So perhaps not only is the quest for an optimal jurisdiction size futile, the very question of optimal jurisdiction size may be misguided.

## REFERENCES

Allers, M.A., & Geertsema, J.B. (2016). The effects of local government amalgamation on public spending, taxation, and service levels: Evidence from 15 years of municipal consolidation. *Journal of Regional Science* 56(4): 659–682.

Andersen, S.C. & Serritzlew, S. (2007). The unintended effects of private school competition. *Journal of Public Administration Research and Theory* 17(2): 335–356.

Andersen, L.B., Boesen, A. & Pedersen, L.H. (2016). Performance in public organizations: Clarifying the conceptual space. *Public Administration Review* 76(6): 852–862.

Andersen, L.B., Heinesen, E. & Pedersen, L.H. (2014). How does public service motivation among teachers affect student performance in schools? *Journal of Public Administration Research and Theory* 24(3): 651–671.

Andrews, M., Duncombe, W. & Yinger, J. (2002). Revisiting economies of size in American education: Are we any closer to a consensus? *Economics of Education Review* 21(3): 245–262.

Andrews, R., Boyne, G.A. & Walker, R.M. (2006). Subjective and objective measures of organizational performance: An empirical exploration. In G.A. Boyne et al. (eds.), *Public Service Performance: Perspectives on Measurement and Management*. Cambridge: Cambridge University Press, 14–34.



Baekgaard et al. 2014

Baldersheim, H. & Rose, L.E. (eds.). (2010). *Territorial choice: The politics of boundaries and borders*. Houndsmills: Palgrave.

Berry, C.R. & West, M.R. (2010). Growing pains: The school consolidation movement and student outcomes. *Journal of Law, Economics, and Organization* 26(1): 1–29.

Bhatti, Y. & Hansen, K.M. (2011). Who marries whom? The influence of societal connectedness, economic and political homogeneity, and population size on jurisdictional consolidations. *European Journal of Political Research* 50(2): 212–238.

Bhatti, Y., Houlberg, K., Thranholm, E. & Pedersen, L.H. (2017). *Kommunalpolitisk barometer 2017*. Copenhagen: VIVE.

Blom-Hansen, J., Houlberg, K. & Serritzlew, S. (2014). Size, democracy, and the economic costs of running the political system. *American Journal of Political Science* 58(4): 790–803.

Blom-Hansen, J., Houlberg, K., Serritzlew, S. & Treisman, D. (2016). Jurisdiction size and local government policy expenditure: Assessing the effect of municipal amalgamation. *American Political Science Review* 110(4): 812–831.

Boyne, G.A. (1996). Scale, performance, and the New Public Management: An empirical analysis of local authority services. *Journal of Management Studies* 33(6): 809–826.

Boyne, G.A. (1998). *Public choice theory and local government: A comparative analysis of the UK and the USA*. Houndsmills: MacMillan.

Boyne, G.A. (2002). Concepts and indicators of local authority performance: An evaluation of the statutory frameworks in England and Wales. *Public Money & Management* 22(2): 17–24.

Boyne, G.A. (2003a). What is public service improvement? *Public Administration* 81: 211–227.

Boyne, G.A. (2003b). Sources of public service improvement: A critical review and research agenda. *Journal of Public Administration Research and Theory* 13(3): 367–394.

Boyne, G.A., Farrell, C., Law, J., Powell, M. & Walker, R.M. (2003). *Evaluating public management reforms: Principles and practice*. Buckingham: Open University Press.

Boyne, G.A., & Walker, R.M. (2005). Determinants of performance in public organizations. Symposium volume of *Journal of Public Administration Research and Theory* 15: 483–639.

Boyne, G.A., Meier, K.J., O'Toole Jr., L.J. & Walker, R.M. (eds.). (2006). *Public service performance: Perspectives on measurement and management*. Cambridge: Cambridge University Press.

Byrnes, J. & Dollery, B. (2002). Do economies of scale exist in Australian local government? A review of research evidence. *Urban Policy and Research* 20: 391–414.

- Cobban, T.W. (2017). *Bigger is better: Reducing the cost of local administration by increasing jurisdiction size in Ontario, Canada, 1995–2010*. *Urban Affairs Review*.
- Cyert, R.M. & March, J.G. (1992 [1963]). *A behavioral theory of the firm*. Oxford: Blackwell.
- Dahl, R.A. & Tufte, E.R. (1973). *Size and democracy*. Stanford: Stanford University Press.
- Denters, B., Goldsmith, M., Ladner, A., Mouritzen, P.E. & Rose, L. (2014). *Size and local democracy*. Cheltenham: Edward Elgar.
- EVA. (2019). *Evaluering af 7-trins-skalaen*. Copenhagen: Ministry of Education.
- Foged, S.K. (2016). The relationship between population size and contracting out public services: Evidence from a quasi-experiment in Danish municipalities. *Urban Affairs Review* 52(3): 348–390.
- Fox, W.F. (1981). Reviewing economies of size in education. *Journal of Education Finance* 6(3): 273–296.
- Fox, W.F. & Gurley, T. (2006). *Will consolidation improve sub-national governments?* World Bank Policy Research Working Paper 3913.
- Gerring, J. (2007). *Case study research: Principles and practices*. Cambridge: Cambridge University Press.

Gordon, N. & Knight, B. (2008). The effects of school district consolidation on educational cost and quality. *Public Finance Review* 36(4): 408–430.

Heinesen, E. (2004). Determinants of local public school expenditure: A dynamic panel data model. *Regional Science and Urban Economics* 34(4): 429–453.

Heinesen, E. (2005). School district size and student educational attainment: Evidence from Denmark. *Economics of Education Review* 24(6): 677–689.

Hirsch, W.Z. (1959). Expenditure implications of metropolitan growth and consolidation. *Review of Economics and Statistics* 41(3): 232–241.

Hooghe, L. & Marks, G. (2009). Does efficiency shape the territorial structure of government? *Annual Review of Political Science* 12: 225–241.

Houlberg, K., Olsen, A.L. & Pedersen, L.H. (2016). Spending and cutting are two different worlds: Experimental evidence from Danish local councils. *Local Government Studies* 42(5): 821–841.

Imbeau, L.M., Pétry, F. & Lamari, M. (2001). Left-right party ideology and government policies: A meta-analysis. *European Journal of Political Research* 40(1): 1-29.

Jugl, M. (2019). Finding the golden mean: Country size and the performance of national bureaucracies. *Journal of Public Administration Research and Theory* 29(1): 118–132.

Kersting, N. & Vetter, A. (eds.). (2003). *Reforming local government in Europe*. Opladen: Leske & Budrich.

Knott, J.H. & Miller, G.J. (1987). *Reforming bureaucracy: The politics of institutional choice*. Englewood Cliffs, NJ: Prentice-Hall.

Koretz, D. (2008). *Measuring up. What educational testing really tells us*. Cambridge, Ma: Harvard University Press.

Lassen, D.D. & Serritzlew, S. (2011). Jurisdiction size and local democracy: Evidence on internal political efficacy from large-scale municipal reform. *American Political Science Review* 105(2): 238–258.

Meier, K.J. & O'Toole Jr., L.J. (2006). *Bureaucracy in a democratic state: A governance perspective*. Baltimore: John Hopkins University Press.

Meier, K.J. & O'Toole, L.J. (2013). Subjective organizational performance and measurement error: Common source bias and spurious relationships. *Journal of Public Administration Research and Theory* 23: 429–456.

Mohr, L.B. (1973). The concept of organizational goal. *American Political Science Review* 67: 470–481.

Mouritzen, P.E. (ed.). (2006). *Stort er godt. Otte fortællinger om tilblivelsen af de nye kommuner*. Odense: University Press of Southern Denmark.

Oates, W.E. (1972). *Fiscal federalism*. New York: Harcourt Brace Jovanovich.

OECD. (2016). *Education at a glance 2016: OECD indicators*, Paris: OECD Publishing.

Ostrom, E. (1972). Metropolitan reform: Propositions derived from two traditions. *Social Science Quarterly* 53(3): 474–493.

Pollitt, C. (2013). The logics of performance management. *Evaluation* 19(4): 346–363.

Ravitch, D. (2013). *Reign of error. The hoax of the privatization movement and the danger to America's public schools*. New York: Alrede A. Knopf.

Sandsør, A.M.J., Falch, T. & Strøm, B. (2017). Local government mergers and school quality. Working paper.

Sawyer, M.C. (1991). *The economics of industries and firms: Theories, evidence and policy*. London: Routledge.

Tiebout, C.M. (1956). A pure theory of local expenditure. *Journal of Political Economy* 64: 416–424.

Treisman, D. (2007). *The architecture of government: Rethinking political decentralization*. Cambridge: Cambridge University Press.

Van Dooren, W., Bouckaert, G. & Halligan, J. (2015). *Performance management in the public sector* (2nd ed.). London: Routledge.

Walker, R.M. & Andrews, R. (2015). Local government management and performance: A review of evidence. *Journal of Public Administration Research and Theory* 25(1): 101–133.

Walker, R.M., Boyne, G.A. & Brewer, G.A. (2010). (eds.). *Public management and performance: Research directions*. New York : Cambridge University Press.

Walker, R.M., Lee, M.J., James, O. & Ho, S.M.Y. (2018). Analyzing the complexity of performance information use. *Public Administration Review* 78(6): 852–863.

Walsh, K. (1991). Quality and public services. *Public Administration* 69(4): 503–514.

Warren, M.E. (2011). Voting with your feet: Exit-based empowerment in democratic theory. *American Political Science Review* 105(4): 683–701.

Wooldridge, J.M. (2009). *Introductory econometrics: A modern approach*. Michigan: South-Western Cengage Learning.

## Tables and figures

Table 1. Single-Year Estimates of Scale Effectiveness in Public Schooling

	Scale economy	Scale efficiency	Scale effectiveness		
	(1)	(2)	(3)	(4)	(5)
	Costs per student in public schools	Number of teaching lessons	Grades final exam	Grades final exam – controlled for individual SES	Share attending upper secondary education
Amalgamated	1,991.84 (1,254.65)	-20.49 (32.09)	0.04 (0.11)	0.02 (0.06)	-0.32 (0.79)
<b>DiD estimators</b>					
Amalgamated * 2004	716.37 (599.60)	14.11 (27.61)	-0.03 (0.06)	-0.02 (0.05)	-0.33 (0.28)
Amalgamated * 2005	329.02 (759.19)	42.24* (23.80)	-0.04 (0.06)	-0.05 (0.05)	-0.49 (0.42)
Amalgamated * 2006	193.00 (1,051.31)	58.97 (36.07)	-0.07 (0.06)	-0.00 (0.05)	-1.12** (0.55)
Amalgamated * 2007	-73.11 (1,195.95)	46.77 (41.29)	-0.07 (0.06)	0.06 (0.05)	-1.33** (0.56)
Amalgamated * 2008	-542.79 (1,284.82)	54.10 (37.23)	-0.16** (0.07)	-0.02 (0.06)	-1.41** (0.54)
Amalgamated * 2009	-872.92 (1,184.44)	43.80 (35.18)	-0.08 (0.07)	0.06 (0.05)	-0.27 (0.53)
Amalgamated * 2010	-854.14 (1,179.61)	29.97 (37.67)	-0.12* (0.07)	0.03 (0.06)	0.14 (0.63)
Amalgamated * 2011	-877.20 (1,181.46)	11.33 (39.36)	-0.16** (0.07)	0.03 (0.06)	-0.64 (0.65)
Amalgamated * 2012	-539.24 (1,357.22)	-9.39 (44.30)	-0.15** (0.07)	0.04 (0.05)	0.36 (0.73)
Amalgamated * 2013	-137.61 (1,361.70)	-4.45 (48.61)	-0.03 (0.08)	0.06 (0.07)	-0.31 (0.66)
Amalgamated * 2014	-84.75 (1,547.43)		-0.08 (0.09)	0.03 (0.06)	0.11 (0.65)
Amalgamated * 2015	-1,156.04 (1,709.71)		0.08 (0.09)	0.09 (0.06)	-0.01 (0.69)
Amalgamated * 2016	-1,045.00 (1,828.29)		0.03 (0.09)		
<b>Control variables</b>					
Dispersal of Settlement	994.96	-65.52***	- 0.29***	-0.00	-1.82***



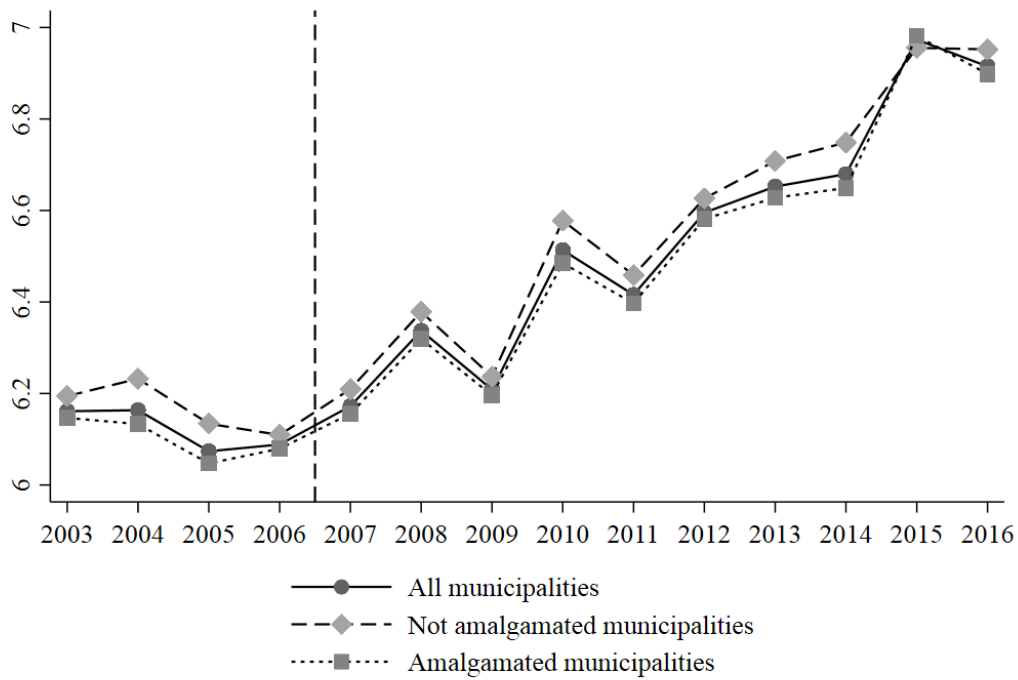
	(666.67)	(13.58)	(0.04)	(0.02)	(0.35)
Fiscal Pressure	-461.80***	-4.34***	-0.01*	0.01*	-0.03
	(76.20)	(1.52)	(0.01)	(0.00)	(0.05)
Social Housing	-101.10*	-2.81***	0.00		0.00
	(59.79)	(1.06)	(0.00)		(0.03)
Share of Students in Single-Parent Families	251.11	2.72	-		-0.58***
			0.06***		
	(164.09)	(3.93)	(0.01)		(0.09)
Share of immigrant students	441.79***	-1.65	-		-0.23***
			0.04***		
	(102.84)	(1.97)	(0.01)		(0.05)
Change in no. of students (2003=100)	-276.74***	-1.86	0.01**	0.00*	-0.02
	(51.58)	(1.29)	(0.00)	(0.00)	(0.02)
Share of socialist seats	83.28**	0.83	-0.00	-0.00***	-0.07***
	(33.92)	(0.76)	(0.00)	(0.00)	(0.02)
<b>Year dummies (ref = 2003)</b>					
2004	-364.73	-11.10	0.06	0.02	0.79***
	(578.48)	(26.10)	(0.05)	(0.04)	(0.26)
2005	392.74	-31.87	-0.01	0.04	0.32
	(696.40)	(22.69)	(0.06)	(0.04)	(0.40)
2006	476.74	-20.85	0.08	0.00	0.37
	(1,003.72)	(34.44)	(0.06)	(0.04)	(0.51)
2007	-628.69	-26.46	0.21***	-0.04	0.18
	(1,188.13)	(40.37)	(0.06)	(0.04)	(0.57)
2008	-281.06	-40.39	0.47***	0.02	0.43
	(1,237.81)	(36.07)	(0.07)	(0.05)	(0.53)
2009	190.85	-35.87	0.29***	-0.03	2.17***
	(1,099.71)	(33.97)	(0.07)	(0.05)	(0.51)
2010	-2,469.77**	-47.20	0.67***	0.03	3.58***
	(1,042.82)	(37.48)	(0.07)	(0.06)	(0.61)
2011	-4,959.30***	-38.22	0.62***	0.02	6.37***
	(1,001.41)	(38.89)	(0.06)	(0.06)	(0.59)
2012	-6,359.90***	-24.16	0.84***	0.00	6.41***
	(1,168.27)	(42.23)	(0.07)	(0.04)	(0.66)
2013	-8,470.35***	-30.48	0.79***	-0.01	7.69***
	(1,175.65)	(46.33)	(0.08)	(0.07)	(0.66)
2014	-4,603.80***		0.88***	0.00	7.34***
	(1,331.78)		(0.08)	(0.05)	(0.63)
2015	-1,730.35		1.10***	-0.04	7.11***
	(1,507.37)		(0.08)	(0.05)	(0.58)
2016	-2,708.00		1.11***		
	(1,659.19)		(0.08)		
Constant	126,105.87***	2,251.09***	8.78***	-0.80**	100.72***
	(10,151.57)	(226.21)	(0.71)	(0.36)	(5.31)
Observations	1,260	990	1,260	1,170	1,170
Adj. R <sup>2</sup>	0.492	0.322	0.719	0.084	0.748

Robust standard errors in parentheses (clustered at each municipality)

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

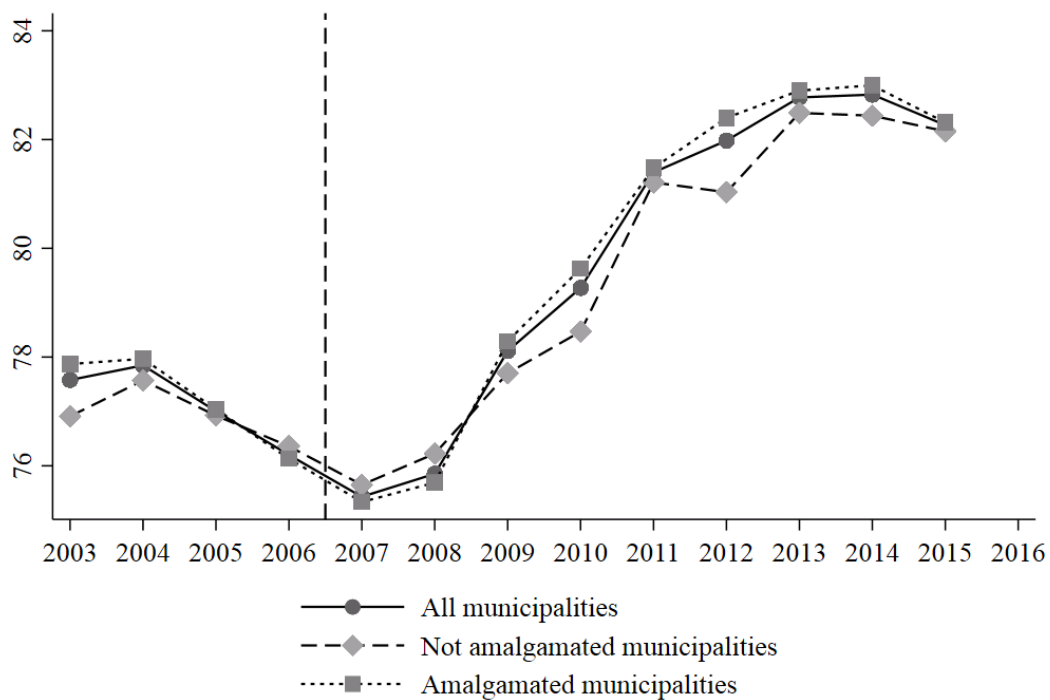
**Figure 1. Group Means on Dependent Variables, by Year.**

Figure 1a. Average Grades on Final Exams



Note: Copenhagen, Frederiksberg and Bornholm omitted.

Figure 1b. Share of 18–21-year-olds Attending and Having Completed an Upper Secondary Education



Note: Copenhagen, Frederiksberg and Bornholm omitted.