Zelda Brutti

Cities drifting apart: Heterogeneous outcomes of decentralizing public education

Abstract

Looking at the decentralized provision of public education in a middle-income country, this paper estimates the impact of local autonomy on service quality, finding large heterogeneity in the effect across different levels of local development. In the year 2002, Colombian municipalities were entrusted with autonomous management of their local public education based solely on a population threshold. I estimate the impact that autonomy has had on education performance across the territory, using a municipality and time fixed-effects model. I find a quality gap arising between highly developed and low-developed autonomous municipalities, in a trend that reinforces over time: the reform has induced regional inequality in education quality. I am able to support the hypothesis that autonomous and nonautonomous municipalities were on similar performance trends before decentralization was implemented, even when looking within different local development ranges. Based on the analysis of detailed municipal balance sheet data and administration indicators, I argue that local administration capacity represents the most likely explanation of why the autonomy-related discrepancies have been arising.

Current version: February 29, 2020
Keywords: decentralization, education inequality, development inequality
JEL-codes: I24, I25, H75, H77
Corresponding author: Zelda Brutti
zelda.brutti@ub.edu

IEB - Institut d'Economia de Barcelona, UB School of Economics, University of Barcelona, Barcelona, Spain.

Cite as: Brutti. *IZA Journal of Labor Economics* (2020) 9:3.
https://doi.org/10.2478/izajole-2020-0003

1 Introduction

Decentralization of public service provision has been at the top of policy agendas in numerous countries over the past decades, involving services such as education, health, public transport, energy supply, and water and sewerage systems. Developing and middle-income countries, in particular, have been transferring responsibilities from a central or regional level down to the municipal one.[1] Such reforms are typically expected to increase welfare through better local preference matching (Musgrave, 1959; Oates, 1972), higher monitoring and accountability (Crook and Manor, 1998; Manor, 1999; Blair, 2000), and more efficient service delivery. However, welfare losses have been shown to result at times, due to inadequate management skills at the very local level, increases in administrative and coordination costs (Breton and Scott, 1978; Panizza, 2004), and corruption among local bureaucrats or resource capture by local elites (Bardhan and Mookherjee, 2000, 2002, 2005, 2006). Positive and negative repercussions from service decentralization may materialize in different proportions across different geographical regions, potentially exacerbating subnational inequalities. In this paper, I show that entrusting Colombian municipalities with autonomy over public education has yielded heterogeneous results in terms of local educational quality, depending on the level of municipal development at the time of responsibility takeover. Education quality in highly developed and low-developed municipalities has been drifting apart over the 10 years following the decentralization reform, as a consequence of autonomy being transferred to the very local level.

This empirical analysis benefits through an unusually clean and simple decentralization criterion: autonomy over the education service was assigned to cities solely depending on whether they exceeded the threshold of 100,000 inhabitants in 2002. This criterion rules out some typical obstacles to the identification of the effects of higher local autonomy: endogenous selection into autonomy by a selected group of local authorities, or the impossibility of disentangling the decentralization process from nationwide changes occurring at the same time. The provision of public education in Colombia is a case worth analyzing, as it represents an instance of decentralization limited to a subset of local authorities, but without performance-driven selection into it: this paper is able to offer new insights on the service decentralization debate benefiting from a quasiexperimental setup. Moreover, this paper isolates the effects of devolving decision powers without the interference of contemporaneous fiscal or political changes – which often come along with decentralization processes. In fact, the Colombian reform transferred managerial responsibility to the local level but left local taxation powers and local representation unaltered.

The empirical strategy adopted in order to isolate the effects of local autonomy is a municipal and year fixed-effects model, which is able to account for permanent differences between local authorities and for nationwide changes over time. I verify that, before the decentralization reform, the performance trends of municipalities which would become autonomous in 2002 were comparable to those of their nontreated counterparts – even when looking within specific development ranges. Identification of the effect through a regression discontinuity (RD) design would, in principle, be suitable to the reform context as well, but it has been discarded due to the sample size being too modest for such a data-intensive strategy.

1 Recent examples are the experiences of Chile, Argentina, Bolivia, Brazil, and Colombia in Latin America; India, Thailand, Vietnam, and the Philippines in Southeast Asia; South Africa, Senegal, Ethiopia, and Uganda in Africa; Ukraine, Serbia, and Bulgaria in Eastern Europe.

Using 13-year-long panel data on standardized student test scores, which proxy education quality, I show that higher autonomy has proven beneficial for highly developed municipalities, but it had detrimental effects for less-developed municipalities. The development-related performance differences grow stronger over time, indicating a gradual manifestation of the consequences of the change in management. The quality differences appear to affect the intensive margin of education quality, since they are not explained by changes in the size or socioeconomic composition of the student pool.

I explore the channels at work behind the arising gap in test scores, by studying the detailed municipal education expenditure data and a set of administration quality indicators. Due to the unavailability of pre-reform data, this part of the analysis is not causal but merely suggestive. Nevertheless, its results are helpful toward building a data-driven opinion on the mechanisms underlying the effects on education quality. Figures suggest that within-country heterogeneity in local administration capacity may have played an important role in explaining heterogeneity in outcomes. A contribution due to differences in local financial resources is not precisely identified but cannot be excluded.

The findings of this paper speak to the effects of large-scale administrative decentralization reforms and to their potentially heterogeneous effects in contexts characterized by significant subnational diversity. This message represents a relevant reference for future public service decentralization reforms, especially if planned in low- and middle-income contexts.

2 Selected Literature

Heterogeneity in the effects of decentralization is modeled in work by Bardhan and Mookherjee (2000, 2002, 2005, 2006), who show how the combination of strong local elites and weak local institutions can cause decentralization to yield underprovision of services to the local poor. Further channels for diversity of impacts across places and people are illustrated in the reviews by Kaiser (2006) and, with a special focus on developing countries, Juetting et al. (2005). These reviews and the vast majority of empirical literature fail to establish any clear link between decentralization and poverty reduction, in addition to documenting higher advantages for the rich with respect to the poor in decentralized contexts.

Focusing on literature related to the key message of this paper – heterogeneity in the impact of decentralization across local development levels – we find studies describing correlations between indicators of local welfare and the spending decisions of local politicians, but without establishing causal relationships between the two. Reinikka and Svensson (2004) find that decentralized school grants in Uganda were subject to local elite capture, but less so in better-off communities. Local governments are found to be more responsive to citizen's needs when the electorate is more informed and when better institutions are in place, in studies by Besley and Burgess (2002) on India and Ferraz and Finan (2011) on Brazil. Faguet and Sanchez (2008, 2014) focus on local financial independence: they look at Colombian municipalities' balance sheet data and establish negative associations between dependence on central government transfers and expenditure on education, but a positive association with public and private school enrollment rates.

Turning the attention to causal analyses, there are a number of studies that aim at isolating the effects of *fiscal* decentralization processes at different levels of local wealth.

These studies address reforms that mostly changed the allocation of *taxing power* across different levels of government. Hammond and Tosun (2011) apply a spatial error model in the US and find that fiscal decentralization, as proxied by government fragmentation, led to gains in employment and economic growth for metropolitan counties, but insignificant to negative impacts for nonmetropolitan ones. The fixed-effects analysis by Zhang (2006) shows that fiscal decentralization in China has promoted (i) regional inequality – mainly due to inequalities in tax bases and, thus, in fiscal burden and (ii) unequal development of nonfarm activities between jurisdictions. Contrary to mainstream findings, Faguet (2004) finds that after a large fiscal and political decentralization process, the poor and marginalized communities of Bolivia took advantage and adapted their expenditure structure to local needs.

To the best of my knowledge, there are very few studies analyzing the effects of *administrative* decentralization, sometimes known as devolution, in combination with heterogeneous local development levels. The paper by Galiani et al. (2008) is close to mine in terms of both type of reform and outcomes studied: the authors show that transferring a number of Argentinian schools from a central to a provincial management yielded positive results in terms of test scores only for schools located in rich municipalities. However, their study on Argentina differs from mine on several aspects. First, this paper benefits from the availability of a quasiexperimental setup in the Colombian case, due to the clear-cut population-threshold rule applied, while, in Argentina, the transfer schedule was the result of negotiations between the federal government and each region. Second, the scale of the reforms analyzed is very different: in Colombia, the whole public education service was transferred to the management of local authorities that had never before held that responsibility, while, in Argentina, an *additional* set of schools was transferred to regions that had managed the same task for over a century. Finally, this paper looks at the results of decentralization to the municipal level, while Galiani et al. (2008) study a case of regional-level decentralization. In sum, it is hard to predict the decentralization outcomes of a reform like the Colombian one, taking the Argentinian experience as a reference. Close in terms of context studied is the working paper by Cortés (2010), who also focuses on the 2001 Colombian reform looking at enrollment outcomes for 2002–2006, but he does not focus on heterogeneity across municipalities. He applies a RD strategy and finds that municipalities that gained education autonomy, on average, increased enrollments of publicly subsidized pupils into private schools. I refer the reader to Section 3.2.2 for a discussion on why the RD strategy might be suboptimal in this context. Finally, the educational outcomes of the 2001 Colombian reform are descriptively explored in the Colombian Central Bank report by Lonzano et al. (2007), who conclude that, overall, the postreform years have witnessed progress in attendance rates but disappointing results in terms of quality and efficacy. In summary, this paper makes a solid contribution to the still-scarce empirical literature on administrative decentralization outcomes, with a special focus on local heterogeneity in the effect, leveraging a particularly research-friendly reform context.

3 Decentralization in Colombia and the 2001 Reform

Starting in the 1980s, Colombia has been undergoing a progressive decentralization process involving political governance, fiscal structure, and the delivery of public services; various authors have looked at the outcomes of this gradual process, some in a qualitative and some

in a quantitative fashion.[2] The reform approved in 2001 left the political and fiscal scenarios unchanged and enforced administrative decentralization,[3] reallocating subnational responsibilities on the delivery of public services.

3.1 Pre-reform Context and Reform Motivations

Colombia is structured into local authorities as follows: there are 32 departments,[4] 1,118 municipalities located within departments, and four special districts (see maps in Figure A1 in Appendix). Local authorities enjoy decision and spending autonomy over a wide range of matters, although the necessary financial resources mainly consist of central government transfers derived from national tax revenues.[5] Central government transfers have historically been accounting for around 90% of the total education expenditure – on average, nationwide – and the remaining 10% is contributed by local authorities, with some local variability in these figures (Borjas and Acosta, 2000, p.6; Iregui et al., 2006, p.31; Santa Maria Salamanca et al., 2009, pp.19–20). Until the 2001 reform, the law formally entitled both departments and municipalities to regulate public education, hire personnel, and invest in infrastructure[6] – and the resulting division of responsibilities was not transparent (Borjas and Acosta, 2000). In practice, however, being the direct recipients of the bulk of education transfers, departments were the key players in the education sector – as is discussed in further detail later. The elimination of responsibility overlaps for the sake of accountability was one of the main goals of the 2001 reform; further goals were improving efficiency and reducing waste in the use of public resources, mitigating the yearly fluctuations in financial transfers, and updating some obsolete distribution criteria.[7]

3.2 Reform Content

Regarding the management of public education, the reform Law 715/2001 (Congreso, 2001) yielded the fundamental change of a clear-cut allocation of responsibility over the service to either municipalities or departments. Municipalities that counted ≥100,000 inhabitants in the year 2002 became "certified in education", meaning these became responsible for the public education service on their territories and recipients of the education transfers from the central government. I refer to these as "autonomous municipalities" hereafter. Municipalities with <100,000 inhabitants were "not certified", and their public education was run by the departments they belonged to. I call these "nonautonomous municipalities". The next subsection further clarifies the concept of autonomy and discusses the shift in responsibilities.

2 Focusing on education outcomes, Borjas and Acosta (2000), Vergara and Simpson (2001), and Caballero (2006) comprehensively illustrate the dynamics and descriptive trends of decentralizing the public education system over the 1990s, agreeing on generally undistinguished results.

3 Sometimes, this type of administrative decentralization is labeled as "devolution" in literature, referring to situations in which the activities of subnational units of government are substantially outside the direct control of the central government (Rondinelli et al., 1983).

4 These represent the regional level, equivalent to "states" in the US or "provinces" in Argentina.

5 Colombia is considered among the most administratively decentralized countries in Latin America, but it is fiscally very centralized (Alesina et al., 2000; Toro, 2006).

6 Law 60 / 1993 (Congreso, 1993) (distributing competencies across levels of government and assigning resources accordingly), Law 115 / 1994 - the "Comprehensive Education Act" (Congreso, 1994), and respective follow-up decrees.

7 For the official document motivating the reform, see: "*Exposición de motivos 715 de 2001 Nivel Nacional*", *Congreso de Colombia,* and *Gaceta del Congreso* 294 *de* 2000. For further discussion, see Sarmiento and Vargas (1997); Alesina et al. (2000); Borjas and Acosta (2000); Vergara and Simpson (2001); and the technical report by DNP (2002).

The 40 municipalities certified in 2001 account for around one third of Colombia's population and pupil share; their size ranges from 105,000 to >2 million inhabitants.

The reform arranged for a transition period of 2 years, 2002 and 2003, during which autonomous local authorities took over the school infrastructure, prepared for the effective management of the service with the assistance of departments, and had the opportunity to reorganize staffing plans on their territories. During these 2 years, temporary financial transfer amounts were set, and from 2004 onward, the new transfer system became fully operational. During the transition period, the new city-level management was not operational yet – it is better described as a time of preparation for the upcoming change and, thus, analogous to pre-reform years.

The 2001 decentralization reform affected not only the education service but also the provision of health care and other smaller public services, such as water and sewerage management. Nevertheless, these other services were decentralized in *all* municipalities instead of following any inhabitant cutoff rule, so that the definition of treatment and control groups used in the empirical analysis is indeed exclusive to the education sector. Among other robustness checks, we test our results against the inclusion of health service controls.

3.2.1 **Local authorities' competencies and transfers before and after the reform**

Table 1 summarizes the education competencies of local authorities before and after the 2001 reform, in addition to indicating the percentages of education transfers to which they are entitled.

As illustrated in the table, the reform left the role of the central government unchanged but polarized managerial responsibilities and financial transfers among local authorities. Before decentralization, departments were recipients of the bulk of education transfers, and municipalities were de facto quite restrained in their decisions (*Dirección de Desarrollo Territorial Sostenible* [DDTS], 2004). Strikingly, departments were hiring 85%–90% of all teachers and having the final word on their placements across the territory (*Corte Constitucional*, 1997, par.16).[8] From receiving a narrow share of transfers and being subject to departmental supervision, autonomous municipalities transitioned into a situation of full managerial and financial independence, while nonautonomous ones gave up their already very limited powers to the respective departments.[9] I will regard autonomous municipalities as having been treated by the decentralization reform and the nonautonomous counterparts as having remained untreated, since both the figures and the anecdotal evidence indicate that a truly substantial change in regime has happened for the former group but not for the latter. Nevertheless, Section A3 in Appendix discusses how a violation of this premise would affect the interpretation of empirical results.

The reform also brought an adjustment in the allocation formulas of education resources to local authorities. In broad outlines, up to 2001, the majority of education transfers were

8 Municipalities were responsible for allocating teachers across schools *within* their territory, and hired the remaining 10%–15% who were not on departmental payrolls (Gómez et al., 2001). Departments also had the final word on education proposals by municipalities, as these had to be taken in accordance with departments and under their supervision (Law 60 / 1993).

9 With, currently, only 3% of the total funds still flowing to nonautonomous municipalities, and with predetermined use. These funds need to be spent entirely on school infrastructure and school material, according to departments' directions (*Directiva Ministerial* 2003; DDTS, 2004, p.7; Law 715/2001, Art.16).

Table 1 Education Responsibilities and transfers by level of government

Central government			
Set school curriculum	Set teacher wages	Set general guidelines	Financial transfers to local authorities
Local authorities			
Up to 2002 (Law 60/1993)		**From 2002 onward (Law 715/2001)**	
		Autonomous municipalities	
Transfers:	84% to department 16% to municipality	Transfers:	100% to municipality
Teacher hiring, training and placement School infrastructure and materials School transport and any auxiliary program	Departments and municipalities, under departments' supervision	Teacher hiring, training and placement School infrastructure and materials School transport and any auxiliary program	Municipality only
		Nonautonomous municipalities	
		Transfers:	97% to department 3% to municipality
		Teacher hiring, training and placement School infrastructure and materials School transport and any auxiliary program	Department only (maintenance duties for municipality)

Notes: Author's illustration, based on Laws 60/1993, 115/1994, and 715/2001 (Congreso, 1993, 1994, 2001); Borjas and Acosta (2000); and National Planning Department (*Departamento Nacional de Planeación* [DNP]; 2002). Percentages are author's derivation: pre-reform values are based on 2001 data in DNP (2002, p.16); post-reform data are based on 2004 data in DNP (2004a, 2004b). Percentages for departments include the four special districts.

assigned based on the number and seniority of teachers employed, with some adjustment based on number of inhabitants, local poverty, and administrative efficiency. From 2002 onward, pupil headcount gained importance in the allocation criteria – even though the number of teachers kept playing a key role – with minor adjustments for local poverty and population density, as before. These changes applied to transfers to *all* local authorities: to autonomous cities and to departments and, thus, to the nonautonomous cities under the supervision of the latter. Both before and after the reform, education transfers were to be used exclusively for the education service, administered in separate accounts and, thus, not fungible with respect to the remaining revenues and expenses of the local authority (*Directiva Ministerial* 2003). Section A3 in Appendix discusses how the interpretation of the main results might be affected by the changes in funding formulas.

3.2.2 **The population threshold, and why not an RD analysis**

The population figures that were used for the 2001 reform were released by the National Administrative Department of Statistics (*Departamento Administrativo Nacional de Estadistica* or DANE). The counts were not prepared ad hoc for the reform but issued in 1993 as forward projections, on the occasion of the general census, when decentralizing education was not being discussed yet. Autonomy was assigned to those municipalities that would exceed 100,000 inhabitants according to the projections for the year 2002. The cutoff was sharply

implemented, and no exceptions were made in either direction; the way population figures increased allows us to set aside any potential suspicion of targeted inhabitant count manipulation.

Beyond its use in the 2001 reform, the 100,000-inhabitant cutoff does not play any significant role in Colombia's legislation and it is not used in other matters involving municipal public service provision. The 100,000 figure appears in a municipal classification scheme that is performed every fiscal year by the central government, based on a combination between current inhabitant count and current municipal revenues. In combination with appropriate current revenues, 100,000 inhabitants may represent the lower bound for a "first category" city.[10] This yearly classification scheme is used to set limits to salaries of the mayor, council members, and administrative staff, as well as limits to general administrative expenditures; the changes are minor across category thresholds. The smaller municipalities (categories fourth to sixth) are entitled to special support transfers. It is worth emphasizing that this categorization is updated every year based on current population estimates and therefore does not fully coincide with the cutoff used for the 2001 reform, which was based on 1993 projections.

The sharp population cutoff rule described above, which was used to assign autonomy to cities, probably calls to mind RD as an attractive identification strategy. In fact, a previously circulated version of this paper[11] used RD as its main identification approach and obtained results that are qualitatively and quantitatively analogous to the ones presented here. Nevertheless, the number of cities that acquired autonomy is small, and even fewer of them are close to the autonomy threshold: for example, between 80,000 and 130,000 inhabitants, there are only 11 treated and 19 nontreated cities.[12] Despite a number of robustness checks, the limited sample size makes it hard to prove *continuity* (and not merely balance) of any confounders at the threshold. In fact, state-of-the-art RD checks such as the McCrary (McCrary, 2008) density test are underpowered in such a data-scarce environment and are absent both in the previous version of this paper and in the RD-based working paper by Cortés (2010). Further, the low number of cities close to the policy threshold forces the researcher to include into the analysis cities that are quite different in size.[13] While it is theoretically possible to compensate for the differences between these cities through the inclusion of appropriate control variables, as well as to show that they are similar along the most policy-relevant dimensions, such a setup is fairly distant from the one the RD strategy was designed for. I therefore choose to focus on a more conventional and robust fixed-effects approach.

4 Data

4.1 Test Scores

Colombia has a long running tradition of standardized testing in public schools; *Instituto Colombiano para la Evaluación de la Educación* (ICFES, "Colombian Institute for the

10 Law 136/1994 and Law 617/2000. The seven categories and their relative inhabitant cutoffs are as follows: Special (500,001 or above), First (100,001–500,000), Second (50,001–100,000), Third (30,001–50,000), Fourth (20,001–30,000), Fifth (10,001–20,000), and Sixth (10,000 or below).

11 Published as *Institut d'Economia de Barcelona* (IEB) Working Paper 2016 / 26. Retrievable at these permalinks: http://diposit.ub.edu/dspace/bitstream/2445/112825/1/IEB16-26_Brutti.pdf; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2858974.

12 This is the main sample used in the previous RD-based version of the analysis.

13 The more restricted sample of Cortés (2010) includes cities between 20,000 and 180,000 inhabitants; the full sample includes cities between <10,000 and >500,000 inhabitants.

Promotion of Higher Education") is the government agency in charge of conducting and assessing the tests across the whole country. Widely accepted as a proxy for secondary education quality is the Saber11 examination, administered to all students completing high school.[14] The Saber11 scores range from 0 to 100 and are standardized by subject at the national level. That is, each student's score is informative about his/her position relative to the national average in each of the examined subjects. Individual-level Saber11 test scores are made available by the ICFES for the years 2000–2012, with information about the school and the municipality to which each student belongs, as well as some information on student background. I focus on students' achievement in mathematics, which is easily comparable to international literature and on which the school environment is regarded to have the highest impact (Figlio and Loeb, 2011; Cronin et al., 2005).[15]

4.2 **Municipal Development Measures**

The development level of Colombian municipalities is being evaluated periodically by government agencies: relevant data are collected by the National Statistics Office (DANE), and the summary indicators are calculated by the National Planning Department (DNP). Up to the year 2013, the most informative and widely used indicator on local development was the Municipal Development Index (hereafter, MDI[16]). The MDI ranges from 0 to 100 and expresses a composite measure of municipal development; it considers "social" or "life quality" variables, such as coverage of electricity, water and sewerage systems, adult literacy rates, and poverty ratios; and "financial status" variables, such as per capita tax revenue and public spending, as well as dependency on central government transfers. The higher the index value, the better is the local development. I use the 2001 MDI index to measure the local development of municipalities at the time of the reform. City size and development level are overall positively correlated, but with high variation across the size range. Figure 1 shows the MDI distribution in our sample, distinguishing between municipalities that gained autonomy due to the reform and those that did not. Figure A2 in Appendix shows the overall MDI distribution without distinguishing between the two groups.

4.3 **The Sample**

Districts are four large local authorities in Colombia, whose nature is mixed between departments and municipalities; previously, before 2002, these enjoyed autonomy over education matters, and they are excluded from the analysis.[17] Also excluded are the cities at the extremes of the size distribution, i.e., >500,000 and <10,000 inhabitants. Those categories of cities are administratively and fiscally incomparable to the medium-sized municipalities our analysis focuses upon and which were targeted by the 2001 reform. Finally excluded are two

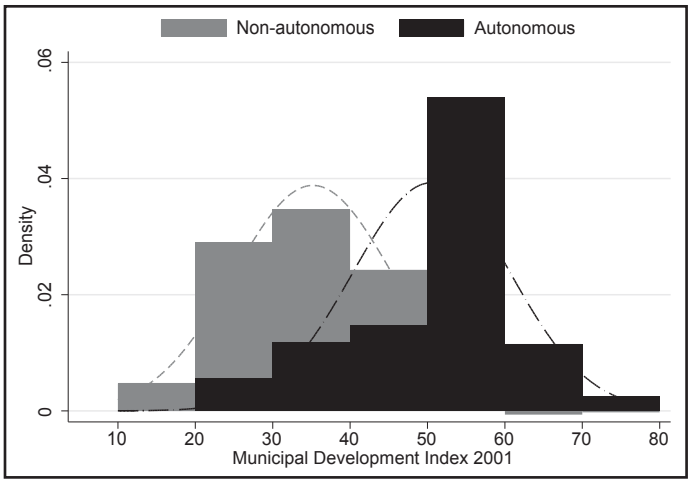
14 That is, students completing 11 years of schooling. The first 9 years are compulsory, and the final 2 years are optional.

15 However, I show and briefly discuss the results for Critical Reading and Sciences as well.

16 Translation from the original *Índice de Desarrollo Municipal* (IDM). Data on the index are provided for public use by the Colombian National Planning Department (*DNP - Departamento Nacional de Planeación*). A new "Overall Performance Index" (*Índice de Desempeño Integral [IDI]*) has been issued starting in 2006 and has now replaced the MDI (2013 onward).

17 Bogotá, Barranquilla, Cartagena, and Santa Marta.

Figure 1 Municipal development in the empirical sample. *Notes:* The figure plots the MDI values for the year 2001 for our sample of cities, distinguishing between those that acquired autonomy in 2002 (darker) and those that did not (lighter). The two distributions overlap in transparency (darkest). Normal curves are overlaid.



municipalities[18] whose freedoms on local education policy had been formally already enhanced back in 1999–2000, even though the substantial implications of the procedure remained unclear. The final sample used is composed of 692 municipalities, out of which 35 became autonomous and 658 remained nonautonomous. Table A1 in Appendix summarizes, for this sample, the descriptive statistics of all variables used in the analysis.

In a robustness check in Section A4 in Appendix, I change the sample by restricting it around the population threshold. This is done in order to alleviate concerns about city size or confounders related to it playing an important role in explaining the main results. The development-related heterogeneity in impact persists, and even strengthens, as I consider cities that are more similar to each other in size.

5 Empirical Framework

The aim is to identify the impact of municipal autonomy over education on student test scores and to pin down any heterogeneous patterns that the effect might display across different levels of local development. Student achievement across the national territory is likely to be co-determined by observable and unobservable factors that are specific to each local area and that might also correlate with city size or local development. A municipality fixed-effects model allows us to account for such factors in a flexible way, without the need for listing them all explicitly.

Let us begin with the following basic fixed-effects model:

$$Y_{it} = \alpha + \tau A_{it} + \boldsymbol{\gamma M}_{i} + \boldsymbol{\delta T}_{t} + \varepsilon_{it} \tag{1}$$

where the test scores in municipality i and year t, Y_{it}, are regressed on the autonomy status A_{it}, which is "0" for all cities before 2002, takes the value "1" in the years from 2002 onward for municipalities that obtained autonomy, and it remains "0" for the others. Vectors of

18 The municipalities of Armenia (Department of Quindio) and San Juan de Pasto (Department of Nariño).

municipality fixed effects $\boldsymbol{M_i}$ and of year fixed effects $\boldsymbol{T_t}$ are controlled for, the latter allowing for the presence of nationwide shocks affecting student performance. The effect of autonomy is captured by τ.

I will call Eq. (1) the "naïve" model, because it can only estimate the *average* effect of autonomy across all independent cities. It does not allow for cities characterized by dissimilar levels of local development to be affected differently by receiving autonomy over education, which is the main hypothesis that this paper wishes to explore. Our baseline model therefore reads as follows:

$$Y_{it} = \alpha + \tau 0 A_{it} + \tau 1 A_{it} \cdot D_i + \gamma \boldsymbol{M_i} + \delta \boldsymbol{T_t} + \varepsilon_{it} \tag{2}$$

where $A_{it} \cdot D_i$ is an interaction between autonomy status and the level of municipal development at the time of autonomy acquisition, so that now the effect of local autonomy is given by $\tau_0 + \tau_1 \cdot D_i$ and depends on development: it represents the average treatment effect for cities that were characterized by development level D_i just before obtaining autonomy.

Achieving identification through our fixed-effects models (1) and (2) relies on the assumptions of linearity in the fixed effects, as well as on the conditional independence assumption $E(Y_{it}{}^0 | M_i, t, A_{it}) = E(Y^0 | M_i, t)$, where Y^0 indicates the quality of education in the absence of autonomy, M_i is the municipality fixed effect, t indicates the year, and A_{it} represents the autonomy status. That is, counterfactual outcomes in the absence of autonomy should be independent of autonomy assignment itself. It is not possible to test conditional independence empirically – nevertheless, it is comforting to observe pre-reform trends in outcomes, which are similar between treated and untreated units; these are shown in Section 6.

A particularly valuable asset in our context is the clear-cut autonomy assignment rule that was used: out of municipal control and a singular event in time. The same is true for the local development measure, which was taken *before* the reform and by an independent governmental agency. These elements characterizing our variables of interest are highly desirable, as they limit the room for omitted variable bias, endogeneity, or reverse causality affecting estimations.

Standard errors are clustered by municipality in all main specifications. The municipal level is the level of treatment by reform design, and we expect most of the correlation in test scores to occur at the municipal level due to the institutional setting. Section A4.3 in Appendix addresses the different ways of computing standard errors.

5.1 Over-Time Evolution

In the baseline model (2), $\tau_0 + \tau_1 \cdot D_i$ estimates the treatment effect averaged across all 10 postreform years (2002–2012). However, it would be reasonable to expect the effects of the reform to appear gradually over time, for at least two reasons: first, cohorts of students taking the Saber11 examination in years further away from the reform have been exposed to the new management for longer[19]; second, cities that have acquired autonomy are likely to require time for (i) the implementation of changes and (ii) their medium- and long-term education plans. Furthermore, recall that under the initial transition period that the reform had arranged (years

19 This is in the same spirit of the exercise performed by Galiani et al. (2008) in their paper on Argentinian school decentralization.

2002 and 2003), the new city-level management was not operational yet, so that those years should be viewed as akin to pre-reform years.

In order to verify whether the expected over-time consolidation of impact is actually corroborated in the data, after estimating the over-time average treatment effect, I run model (2) selecting postreform time periods that are increasingly distant from the year 2001. I predict that τ_0 and τ_1 will yield estimates that grow stronger in absolute values as more distant time periods are used.

6 Results and Discussion

Table 2 shows the main estimation of the consequences of the 2001 decentralization reform on local education quality. Column (1) shows the estimation of the "naïve" model: we can see that on average, emancipation on educational matters was beneficial for cities, as the mean test scores increased by 0.67 points – around 0.3 standard deviations of the municipal average test scores. This overall positive result, however, hides important heterogeneity – as hypothesized – which is unveiled in Column (2), which reports the results of our baseline model. I find a strong interaction between the level of local development at the time of autonomy acquisition and the impact of autonomy itself, and the linear approximation of such interaction is 0.06 points for each additional positive step on the MDI scale. These estimates imply that the average test scores in cities at the lower end of the MDI spectrum are negatively affected by autonomy, while those at the higher end of the spectrum are positively affected. The MDI threshold at which the effect point estimate switches from being negative to positive is around 40. Analyzing the MDI distribution in our sample (plotted in Figure 1; Figure A2 in Appendix), one can observe that two thirds of all municipalities lie below an MDI of 40, and close to 20% of those that acquired autonomy through the 2001 reform lie in this range.

Table 2 continues with Columns (3)–(7), which report the results of the time-progression exercise described in Section 5.1. As expected from intuition, the reform impact and its heterogeneity across local development intensify over time. In the years immediately following the reform (2002–2003), no effects are detected at any MDI level. Starting 2004–2005, statistically significant positive effects appear for highly developed cities with MDI >70, extending to those >50 in the years 2006–2007. In 2008–2009, low-developed cities with MDI <20 start showing statistically significant losses with respect to the pre-reform period and, at the end of our observation window (2010–2012), those <30 also show the same. Using these estimates, I illustrate the over-time escalating heterogeneity in impact as follows. Figure 2 shows the coefficient estimates for τ_0 and τ_1 from the baseline model (2): we can see that over time, the effect of autonomy for low-developed cities becomes increasingly negative (left panel: test score losses for hypothetical cities whose MDI is zero), and the gains associated with higher development become increasingly larger (right panel: test score gains for each additional MDI point). Figure 3 plots the overall effect of autonomy for cities characterized by different levels of development, over time.

The descriptive statistics in Table A1 in Appendix shows that the pre-reform standard deviation of municipal average test scores is 1.52 points, and the postreform standard deviation is 2.30 points, which helps in assessing the proportion of the main results in Table 2.

Table 2 The effect of municipal autonomy on test scores

	Period average		Over-time evolution				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Naive	Baseline	2002–2003	2004–2005	2006–2007	2008–2009	2010–2012
Autonomy	0.67***	−2.53***	−0.67	−1.17*	−1.47**	−2.73***	−5.54***
	(0.21)	(0.87)	(0.62)	(0.67)	(0.70)	(1.00)	(1.44)
Autonomy × MDI		0.06***	0.02	0.03*	0.04***	0.07***	0.14***
		(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	8,734	8,734	2,742	2,665	2,707	2,705	3,367
N groups	692	692	692	692	692	692	692
R-sq.	0.51	0.51	0.51	0.77	0.69	0.44	0.44
Mean *y*	42.34	42.34	41.33	41.54	42.69	41.89	42.41

Notes: SEs clustered by municipalities in parentheses.
FE, fixed effects; SE, standard errors.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

On average, the 2001 reform increased local education quality by 0.44 pre-reform standard deviations (or 0.29 postreform ones) for autonomous cities, but with highly uneven distribution of the impact across development levels. Looking at the extremes of the development distribution, autonomous cities in the highest decile – characterized by a 2001 Development Index >60 – saw average test score increases of 0.70 pre-reform standard deviations (0.47 postreform ones) over the 10 years following the reform. On the other hand, autonomous cities in the lowest development decile – characterized by a 2001 Development Index <33 – experienced test score deterioration of around 0.35 pre-reform standard deviations (0.24 postreform ones).

These effects are large and are characterized, as seen before, by a distinct intensification over time: by the end of the observation period – 10 years after the reform implementation – the impacts reach double the size of the aforementioned period averages. The decentralization reform induced substantial inequality in local education quality, to the favor of highly developed municipalities and to the loss of low-developed ones.

In Appendix Table A2 in Appendix, I also show the estimation results for two additional subjects, which are typically included in international assessment programs: Critical reading and Sciences.[20] While the pattern of findings resembles the one for the results on Mathematics, the impact of the reform on these subjects is weaker and substantially slower in materializing. In both subjects, the baseline specification (2) estimates the development-related gradient to be around one third of what was found for Mathematics. Moreover, the effects of autonomy remain sluggish up until the years 2010–2012, where they eventually become strong and significant but reach only about half (Critical reading) and one third (Sciences) of the magnitudes pertaining to Mathematics. In sum, my results confirm the rather recurrent finding of students' Mathematics performance exhibiting higher policy-related responses with respect to the performance of other subjects (Figlio and Loeb, 2011; Cronin et al., 2005).

20 Both "Critical reading" and "Sciences" were introduced to Saber11 starting in 2014, as a result of a restructuring of the examination (ICFES, 2013); before that year, "Critical reading" was split into "Language" and "Philosophy" components, while "Sciences" was separated into "Physics", "Chemistry", and "Biology". I therefore construct "Critical reading" and "Sciences" scores by taking the arithmetic mean of their 2000–2012 components.

Figure 2 The effect of municipal autonomy on test scores. *Notes:* The figure plots the coefficients on municipal autonomy (left) and on the interaction between autonomy and development (right), estimated at different postreform periods with baseline model (2). Black squares indicate over-time average estimates, and spikes represent 95% confidence intervals. The reform transition period is shaded in gray.

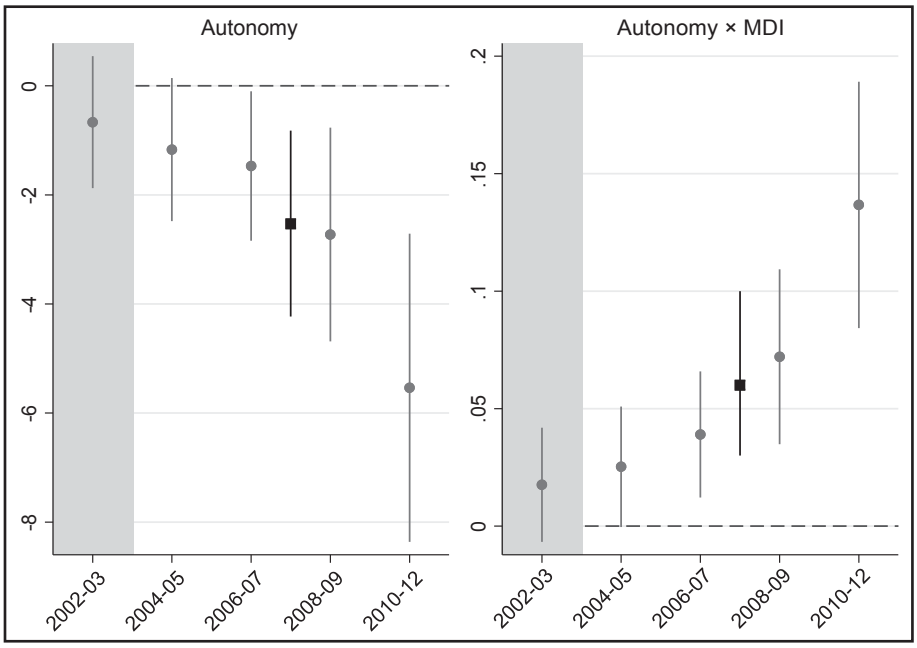
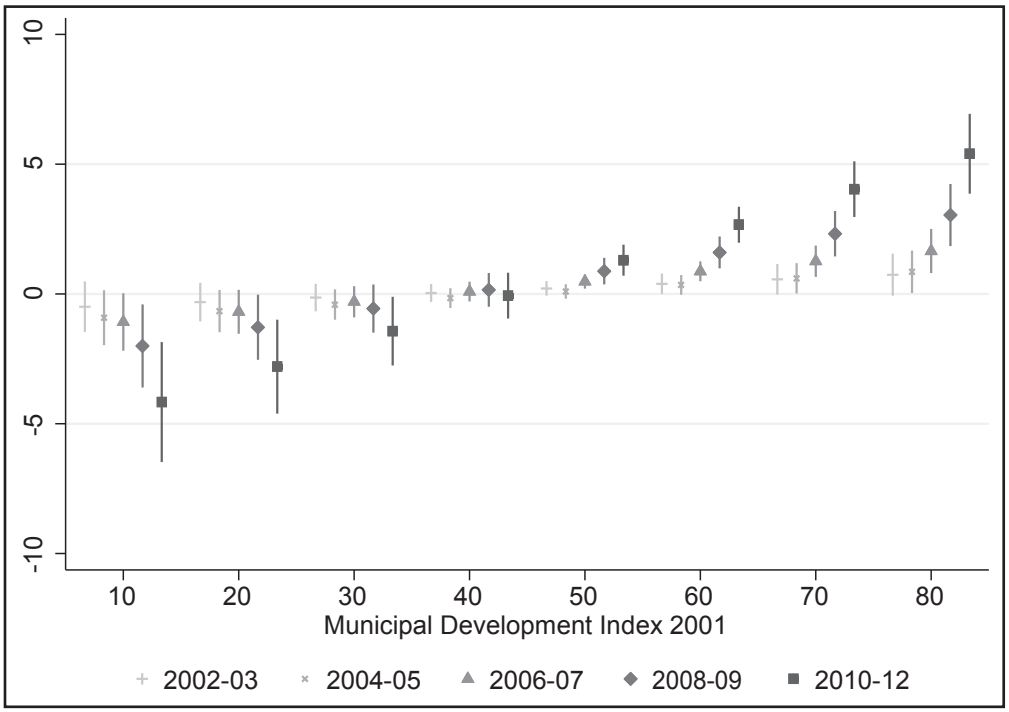


Figure 3 The overall effect of municipal autonomy on test scores, by MDI and time. *Notes:* The figure plots the marginal effects of municipal autonomy on test scores, by level of municipal development (on *x*-axis) and at different postreform periods (see different marker shapes). Spikes represent 95% confidence intervals.



6.1 **Testing Parallel Trends**

While accounting for differences in levels, the fixed-effects strategy used for identification relies on the assumption that counterfactual *trend* behaviors are the same between autonomous and nonautonomous cities. Considering that counterfactual trends are unobservable and thus untestable, literature typically settles for testing the similarity of *pre-reform* trends instead. I use a well-known causality check in the spirit of Granger (1969) and Autor (2003).[21] It consists in ensuring that the treatment effect of the policy arises only *after* the policy has been implemented, and that pre-policy outcomes evolve on parallel trends for treatment and control units, unable to predict treatment. In practice, the original fixed-effects model equations are augmented with interactions between the treatment variable and leads and lags of the year of policy implementation. Interactions representing future policy changes – i.e., those with pre-reform time periods – should not prove significant. If treatment interactions are insignificant in pre-reform years, then treated and nontreated units were moving on similar outcome trends in those years, and it would have been impossible to predict future treatment status based on those trends. The outcomes of this test are shown numerically in Table 3 and graphically in Figure 4. In Table 3, the reform year 2002 is taken as baseline and both columns show the estimates of the remaining *Year* × *Treatment* interactions, which express differences with respect to 2002. Column (1) and the upper panel of Figure 4 report estimations of the "naïve" model (1), in which *Treatment* simply corresponds to *Autonomy*, i.e., having received autonomy in 2002. Beyond the expected growth in impact over time, we can appreciate how the treatment effect indeed arises only in post-reform years and after the "transition period" (2002–2003). No "buildup" to the reform is detected in earlier years – or in other words, pre-reform outcomes do not show any significant interaction with the future autonomy status. Column (2) and the lower panel of Figure 4 report estimations of the baseline model (2), in which *Treatment* corresponds to the interaction *Autonomy* × *MDI*, i.e., having received autonomy in 2002 and being characterized by development level *MDI*. We can see how not even *development-specific* pre-reform outcomes have any significant relationship with future autonomy status. In conclusion, I find support for the hypothesis of similar counterfactual trend behaviors between treated and nontreated cities, even when allowing for development-specific gradients.

In Section A4, I show the results for an alternative method of seeking support for the hypothesis of common pre-reform behavior of treated and nontreated cities, by the level of development. It consists in augmenting the fixed-effects equations with development-specific time trends and check whether this addition makes the treatment redundant in explaining outcomes.[22]

If educational outcomes in cities characterized by different levels of development could be explained by distinct over-time evolution patterns, then autonomy assignment should not matter anymore and our estimates of τ_0 and τ_1 become insignificant in the baseline model (2). I allow for specific trends by high and low development,[23] by development tertiles, quartiles, and quintiles; in all specifications, I find that the impact of autonomy and its development-specific gradient are preserved, although point estimates for both reduce slightly.

21 This discussion is inspired by Angrist and Pischke (2009), Chapter 5.

22 This approach is used, among others, by Besley and Burgess (2004) in their paper on labor regulation in India.

23 Defined as MDI being above or below 40, which is the threshold at which the main results found the reform effect to switch from positive to negative.

Table 3 Leads and lags

	(1)	(2)
	Naive	**Baseline**
y2000 × Aut. (× MDI)	−0.42* (0.25)	−0.01* (0.00)
y2001 × Aut. (× MDI)	−0.33 (0.22)	−0.01* (0.00)
y2003 × Aut. (× MDI)	−0.31* (0.19)	−0.01** (0.00)
y2004 × Aut. (× MDI)	−0.33*** (0.10)	−0.01*** (0.00)
y2005 × Aut. (× MDI)	−0.14 (0.18)	0.00 (0.00)
y2006 × Aut. (× MDI)	0.20 (0.31)	0.00 (0.01)
y2007 × Aut. (× MDI)	0.02 (0.17)	0.00 (0.00)
y2008 × Aut. (× MDI)	0.43** (0.19)	0.01** (0.00)
y2009 × Aut. (× MDI)	0.68*** (0.20)	0.01*** (0.00)
y2010 × Aut. (× MDI)	1.12*** (0.27)	0.03*** (0.00)
y2011 × Aut. (× MDI)	1.11*** (0.38)	0.03*** (0.01)
y2012 × Aut. (× MDI)	0.68*** (0.25)	0.02*** (0.00)
Municipality FE	Yes	Yes
Time dummies	Yes	Yes
N	8,734	8,734
N groups	692	692
R-sq.	0.51	0.51
Mean *y*	42.34	42.34

Notes: SEs clustered by municipalities in parentheses.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

6.2 Illustration: The Evolution in Performance for Two Groups of Cities

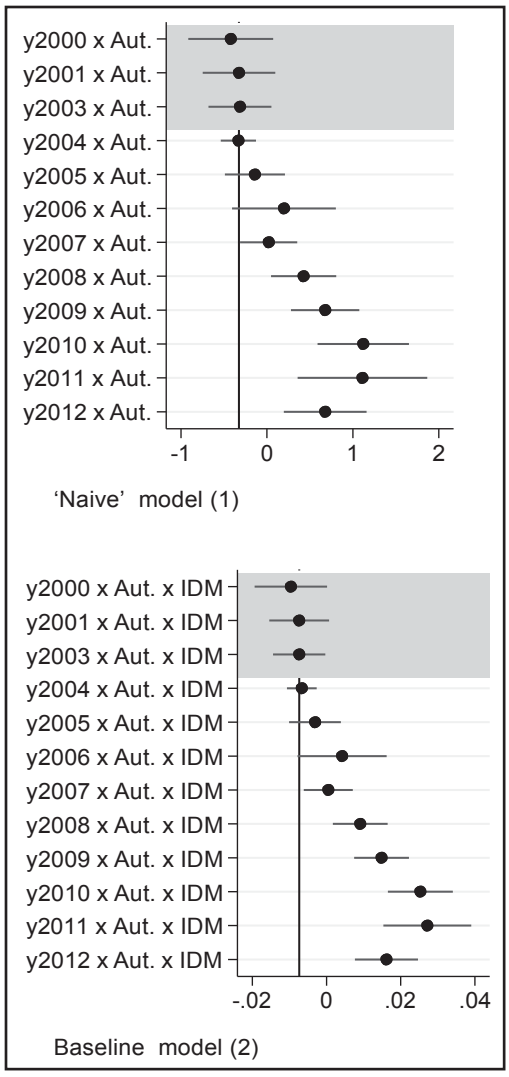
I present an exercise based on the main results so far described, with the purpose of illustrating the over-time evolution in performance for highly developed and low-developed cities – both in absolute terms and relative to each other. It is important to emphasize that the following exercise and figures *do not* reflect the fixed effect identification strategy used to obtain the main results of this paper: instead, the exercise is *based on* those main results, with the purpose of revealing potentially interesting dynamics, which have not been stressed so far.

Based on the results of Table 2, I divide cities into highly developed and low-developed depending on whether their MDI in 2001 was above or below 40: recall that cities with an MDI >40 improved, on average, their performance after obtaining autonomy, while those with MDI <40 declined on average.[24] Figure 5 compare the test score performances of autonomous, non-autonomous, highly developed, and low-developed cities over time – expressed in deviations from the yearly national average. Figure A3 in Appendix shows the evolution in raw scores instead.

Subfigure (a) focuses on the highly developed group and shows the performance trends of autonomous and nonautonomous cities separately. One can observe that the performances were quite similar to each other during the pre-reform years (2000–2001) and during the transition period; then, a performance gap opens up, starting in the mid-2000s, growing wider over time. These patterns had emerged already in our main results. As an additional result, the figure suggests that the most dynamic group of the two is the autonomous one, whose

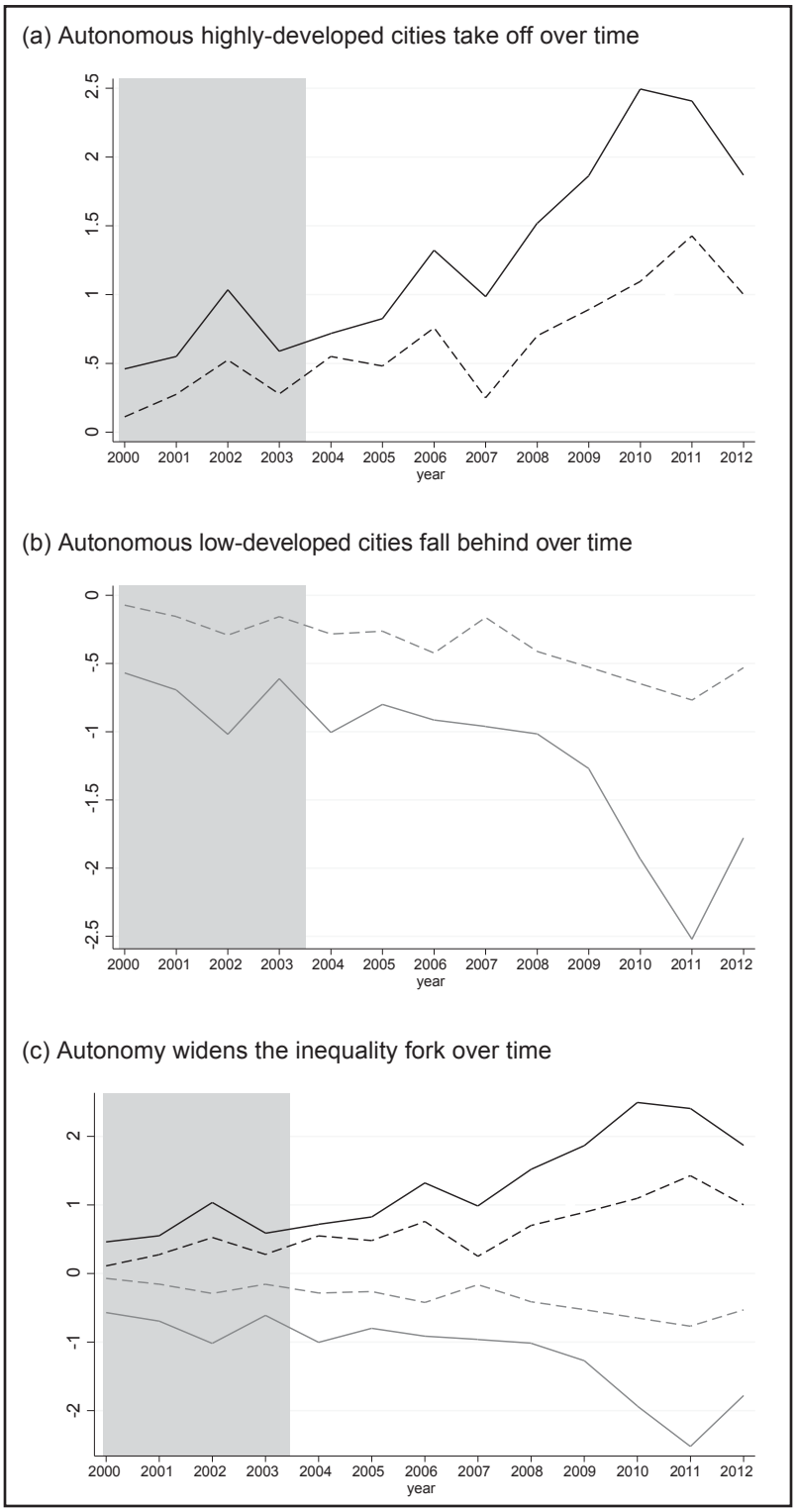
24 From Table 2, Column (2): 2.53 / 0.06 = 42.16, which I round to 40 for clarity and without changes in conclusions.

Figure 4 Leads and lags. *Notes:* The figure plots interactions between autonomy and year dummies ("naive model") and three-way interactions between autonomy, local development, and year dummies (baseline model); point estimates represent differences with respect to 2002. Spikes represent 95% confidence intervals. Pre-reform and transition years are shaded in gray.



performance improvement reveals itself as the driver of the test score divergence during the second half of the decade. Subfigure (b) looks at the low-developed group and again shows the performance of autonomous and nonautonomous cities separately. As for their highly developed counterparts, similar performance between autonomous and nonautonomous cities can be observed before the reform and during early postreform years, while a performance fork slowly opens up around the middle of our observation window. Once again, the driver of the divergence appears to be the autonomous group, as illustrated by its performance decline relative to the national average. Finally, Subfigure (c) plots the four groups of cities against each other. An additional upshot here is that even performance trends between highly developed and low-developed cities were not too distant from each other's in the early 2000s. The subfigure also reveals that a more modest development-related performance gap arises over time even between nonautonomous cities (dashed lines), but the combination with autonomy greatly amplifies it (solid lines).

Figure 5 Evolution of test scores for the two groups of cities over time. (a) Autonomous highly developed cities take off over time. (b) Autonomous low-developed cities fall behind over time. (c) Autonomy widens the inequality fork over time. *Notes:* The figure shows the test score deviations from the national average. Autonomous cities are solid, and nonautonomous ones are dashed. Highly developed cities are in darker color, low-developed ones are in lighter color. Pre-reform and transition years are shaded in gray.



6.3 **Treatment of Nonautonomous Municipalities**

Section 3.2.1 and Table 1 described how the 2001 reform determined an important change of managerial regime for those municipalities that obtained autonomy over education, while those that did not can be considered "untreated". Nevertheless, it may be worth devoting a discussion to the contingency in which nonautonomous municipalities were also in a certain sense treated by the reform – and there are two potential scenarios that come to mind. In the first, let us assume that, contrary to what is suggested by pre-reform reports and figures, municipalities did actually enjoy nonnegligible power on education matters (e.g., through unobserved lobbying actions) and lost it after 2001. In the second scenario, let us assume departments began behaving differently after the largest cities left their responsibility so that nonautonomous cities started to be treated differently by departments as a consequence of the reform.

In both of these cases, our reform estimates would represent a combination of the effect of greater autonomy for the larger cities and of the "other treatment" received by the smaller cities. In fact, we can show this more formally as follows[25]: let

$$Y_{it} = \alpha A_i + \tau_A D_t + \pi_A T_t + \varepsilon_{it}$$

be the outcomes for cities that receive autonomy in 2002 (Type-A cities), where $(\alpha A_i + \pi_A T_t)$ is the expected outcome evolution in the absence of autonomy, D_t is a postreform dummy, and τ_A is the effect of obtaining autonomy; let

$$Y_{it} = \alpha B_i + \tau_B D_t + \pi_B T_t + \varepsilon_{it}$$

be the outcomes for cities that do not receive autonomy in 2002 (Type-B cities), where τ_B is the effect of obtaining the "other treatment". Our parallel trend assumption, tested in Section 6.1, implies that $\pi_A = \pi_B = \pi$. Then we can write as follows:

$$Y_{it} = \alpha_i + \tau_B D_t + (\tau_A - \tau_B)\, D_t A_i + \pi T_t + \varepsilon_{it} \tag{3}$$

where A_i is a dummy for being a Type-A city. Further letting the municipal fixed effects $\alpha_i \equiv \alpha + \gamma M_i$, the time fixed effects $\tau_B D_t + \pi T_t \equiv \delta T_t$,[26] $D_t A_i = A_{it}$, and $\tau_A - \tau_B \equiv \tau$, we are back to the main model (1). Equation (3) shows that the standard fixed-effects analysis I present is in fact estimating $(\tau_A - \tau_B)$, the difference between the impact of greater autonomy and the impact of the "other treatment" received by smaller cities. Similarly, the baseline model (2) is in fact estimating $\tau_0 \equiv \tau_{0A} - \tau_{0B}$ as the main effect of the reform and $\tau_1 \equiv \tau_{1A} - \tau_{1B}$ as the interaction effect between the reform and the development level. In the following paragraphs, I discuss the consequences of the possibility that τ_B, τ_{0B}, $\tau_{1B} \neq 0$.

The case that $\tau_B \neq 0$ would require changing the interpretation of the main results from the effect of a pure autonomy gain to the effect of an "autonomy gap" arising after 2002, in part composed of autonomy gain by larger cities and, in part, autonomy loss experienced by smaller cities. However, this change in interpretation would not affect the main scope and message of this paper – which is illustrating the heterogeneous impacts of local autonomy across the development spectrum. Therefore, the most important question to address is whether the hypothesis τ_{0B}, $\tau_{1B} \neq 0$ might threaten this leading result, explaining or partially explaining the

25 Credit and thanks to an anonymous referee for this formalization.

26 This embodies the classical assumption of no time-varying unobservables affecting the two groups differently.

worse outcomes for low-developed and better outcomes for highly developed areas following autonomy. This question can be answered by discussing three basic options:

(1) after the reform, nonautonomous cities experienced changes, but these were unrelated to their level of development;
(2) after the reform, nonautonomous cities experienced changes, with disproportional gain for highly developed cities and penalization of low-developed ones;
(3) after the reform, nonautonomous cities experienced changes, with disproportional penalization for highly developed cities and gain for low-developed ones.

Option 1 clearly leaves the leading message unaltered. Option 2 would imply that baseline results represent, in fact, a lower bound of the true interaction between autonomy and development. I find that highly developed autonomous cities do better than what we would expect in the absence of autonomy, but if that expectation is based on "control" cities, *which have in fact improved performance too* due to the reform, the true performance growth of the former group would be underestimated. The opposite is true for low-developed cities. On the other hand, Option 3 would imply that the baseline results may in part or entirety be due to development-related changes in performance among nonautonomous cities, rather than due to development-related changes among autonomous ones. This scenario is thus the most problematic one, but also arguably the least plausible. Empirically, I found evidence against this possibility in the illustration exercise of Section 6.2, which shows that nonautonomous highly developed cities continue smoothly on their slightly positive course and low-developed ones on their slightly negative course after the reform, without exhibiting any visible path breaks, which would make one suspect the contingency insinuated in Option 3.

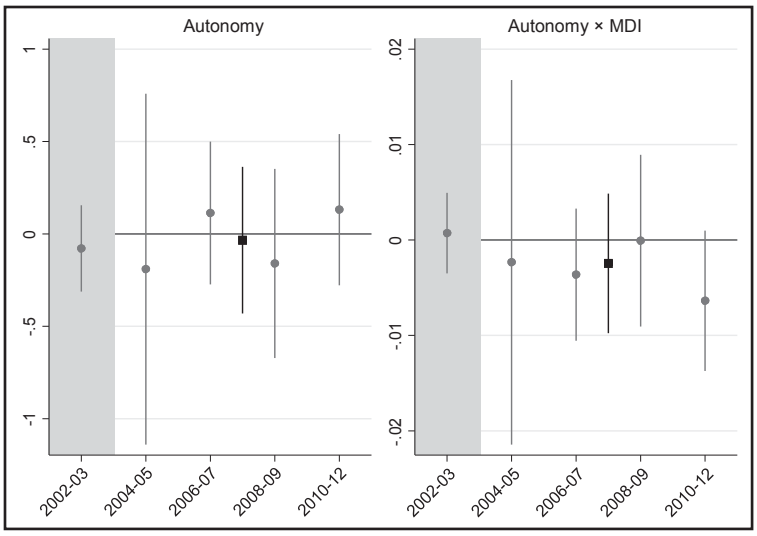
In sum, it is appropriate to recognize the impossibility of fully excluding the chance that cities with <100,000 inhabitants may have experienced changes in their education service after the 2001 reform – even though the analysis of the institutional setting and anecdotal evidence suggest otherwise. In this section, I have discussed how this possibility would affect the results of this paper, and conclude that its main contribution of showing development-contingent outcomes following the decentralization of public education would carry through.

6.4 **Potential Compositional Effects**

One may wonder whether the estimated impacts represent an actual change in local education quality, or whether they might be explained by changes in the pool of test takers, i.e., a compositional effect. A possible conjecture, for instance, could be that low-developed cities used their autonomy to promote rather-inclusive education policies, while the highly developed ones favored upper-class policies instead. This would translate into wider high school participation but lower average grades in the former group, and more-restricted participation but higher average grades in the latter group. In fact, Faguet and Sanchez (2014) claim positive associations of the gradual decentralization process in Colombia with both public and private school enrollment rates over the period 1994–2004, and the previously mentioned RD-based working paper by Cortés (2010) finds positive effects on private enrollment from the 2001 reform.

I check for compositional effects of this sort by estimating the baseline model (2) using the number of test takers and the share of disadvantaged-background test takers as outcomes.

Figure 6 The effect of municipal autonomy on number of test takers (% change). *Notes:* The figure plots the coefficients on municipal autonomy (left) and on the interaction between autonomy and development (right), estimated at different postreform periods. Black squares indicate over-time average estimates, and spikes represent 95% confidence intervals.



The number of test takers is log-transformed, so that the estimated impacts approximate the percentage changes. Disadvantaged-background test takers are those characterized by low socio-economic status, defined by whether the family lives on <2 minimum salaries at the time in which the exam is taken, as self-reported by the student; the information is missing for years 2004–2007.

The results are illustrated in Figures 6 and 7, which are drawn based on the results reported in Tables A3 and A4 in Appendix. I find a 16% decrease in the size of the average test-taker pool of autonomous cities after the reform, but the reduction is not related to development at any point in time and is thus unable to explain the development-dependent pattern in test score outcomes.[27] I do not find any significant impact of the reform on the share of disadvantaged test takers, neither on average nor at any specific point of the development spectrum.

In conclusion, I find no evidence to support the conjecture that the treatment effect on test scores are of compositional nature, and I attribute the differences in findings with respect to the two studies mentioned above to the differences in time periods examined and empirical methods used[28]. Instead, my results are aligned with the findings of Galiani et al. (2008), who claim no compositional effects related to enrollment shifting between the public and private sectors in their study on Argentina. The 2001 reform-induced changes in Colombian test scores are likely to reflect intensive-margin changes in local education quality.

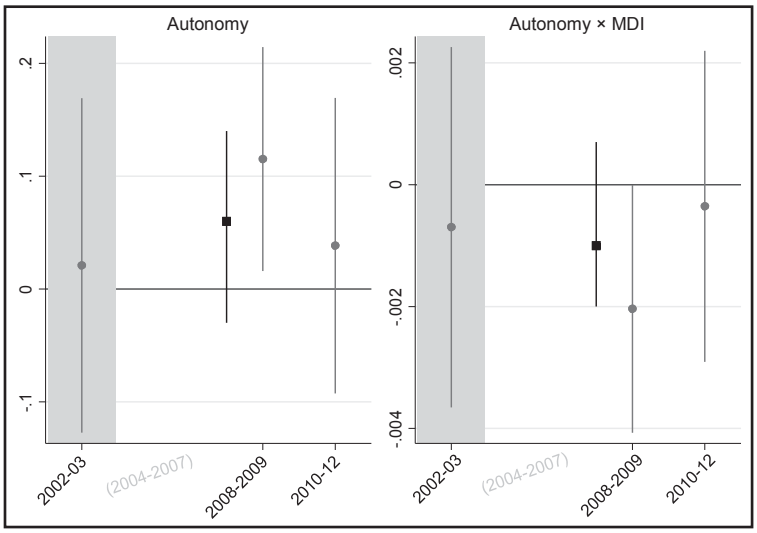
7 Channels

In this section, I offer suggestive evidence that helps in inferring the channels through which the 2001 decentralization reform may have been inducing development-related performance

27 I am not able to pin down any significant changes in the number of test takers from private schools either (not shown).

28 In particular, the 2010 paper uses data on the period 2002–2006 (only 2 years past the reform transition period) and is based on RD identification: see Subsection 3.2.2 for a discussion on why such results may have to be taken with caution. The 2014 paper looks at the decade prior to the reform I analyze and establishes correlations rather than causal impacts between public service outcomes and local financial independence measures.

Figure 7 The effect of municipal autonomy on share of disadvantaged-background test takers. N*otes:* The figure plots the coefficients on municipal autonomy (left) and on the interaction between autonomy and development (right), estimated at different postreform periods. Black squares indicate over-time average estimates, and spikes represent 95% confidence intervals. In the period 2004–2007, socioeconomic status data were not collected.



differences between cities. Based on the literature on service delivery decentralization in developing countries, I explore two classical dimensions that have the potential to explain local performance differences: financial resources and skills (Rondinelli, 1981). By financial resources, I mean the amount of per capita funding that the local authority can rely on and spends on the service – in this case, education. By skills, I broadly mean local administration quality, which I additionally decompose into diligence (application of rules), capacity (amount of human and technological capital available to the administration), and corruption (fraudulent behavior involving the public administration).

Data on both funding and administration quality are available only for the postreform years – some starting in 2002, while others in 2005 or even 2008. Financial data stem from detailed balance sheet reports that municipalities are required to file at the end of each financial year to departments and to the central government.[29] Local administration quality data are collected by the central government on a yearly basis, through a national agency.[30]

Due to the unavailability of pre-reform data and of financial data for nonautonomous cities, it is impossible to aim at producing causal evidence in this section. Comparisons between pre- and postreform behaviors or outcomes are not feasible. The following analysis focuses solely on autonomous cities and provides comparisons between highly developed and low-developed ones: the goal is to use the data to build informed opinions on the likelihood of each channel to have contributed to the policy effects. As in the illustration in Section 6.2, highly developed and low-developed cities are defined according to the MDI threshold emerging from the earlier main results: cities with an MDI >40, which on average gained due to the

29 "*Ejecuciones municipales, formato largo*". Reported yearly by municipalities to the government agency DNP. Source: Universidad de Los Andes, Bogotá, 2015.

30 DNP-DDTS (*Departamento Nacional de Planeación - Dirección de Desarrollo Territorial Sostenible*) is the government agency in charge of the study.

autonomy assignment; and cities with an MDI <40, which on average deteriorated in educational performance.

Table 4 compares the financial and administration indicators of highly developed and low-developed autonomous cities. Due to the small sample size available, only the most clear-cut differences are identified at statistically significant levels. Looking at per-pupil education spending, I am unable to pin down any statistically significant differences between the two groups, neither in the total amount nor in any of the three subcategories (salaries, infrastructure and material, and others). Considering point estimates of the total spending amount, highly developed cities exceed low-developed ones by 8% – which may indicate a nonnegligible difference in resources devoted to education. Still looking at nonprecise point estimates, the "Others" spending category – which includes items such as extracurricular activities, school transport, and additional teacher training programs – appears to be the most unbalanced toward high-development cities, with a 24% larger spending compared to the low- developed counterparts. On the other hand, and somewhat unsurprisingly, spending in infrastructure is approximately 11% higher in low-developed cities. Low-developed cities see, on average, 2.58 fewer students per teacher with respect to highly developed ones.

On the revenue side, there is no significant difference in per-student central government transfers that municipalities receive to finance education services (*SGP Educación*) – and even the point estimate is low, at only 2%. This suggests that, if indeed more resources are spent on education among highly developed cities, these are likely to integrate the dedicated governmental transfers with own financial resources.

Along the dimension of administration quality, the differences between the two groups of cities are more striking. The first indicator I look at measures the "correct application of accounting standards" (*Índice de Cumplimiento de Requisitos Legales*), i.e., the diligence with which the local administration is found to comply with national norms in its use of government transfers. The frequency of detection of accounting irregularities and illicit use of funds determines the rating received on this index (DNP, 2014). I find highly developed municipalities to perform better on average (+11%) than their low-developed counterparts on this evaluation, at a 10% confidence level. What is even more impressive, however, is the large and significant difference between the two groups (+65%) on the second indicator considered, the "administration capacity index" (*Índice de Capacidad Administrativa*), which measures the extent to which the municipal administration appears suitable and prepared to perform its tasks thoroughly. The elements factoring into the index are the stability of managerial employees, the level of education and competency of clerks, the availability of suitable information technology (IT) equipment, and the level of automation of administrative processes (DNP, 2011). Finally, the third indicator analyzed measures corruption, i.e., the instances of criminal sanctions executed against local public administration and its employees each year (*Observatorio Transparencia y Anticorrupción* [OTyA], 2017). Low-developed cities actually report a lower corruption average (–10%) with respect to the highly developed group, even though the difference is not precisely estimated.

In view of the above findings, it is not possible to exclude a contribution of the financial aspect toward explaining the different impacts that autonomy has had on highly developed and low-developed cities. The higher tax capacity of highly developed cities, combined with the new

Table 4 Education finance and administration quality in autonomous cities

	High-D	Low-D	Difference (%)		N	N cities
Education finance						
Total spending	1,175.24 (28.11)	1,088.50 (61.19)	86.74 (66.24)	(+8%)	242	35
Salaries	939.99 (21.00)	861.49 (45.35)	78.50 (49.41)	(+9%)		
Infrastructure and materials	97.38 (6.44)	109.75 (13.66)	−12.37 (15.10)	(−11%)		
Others	87.18 (6.77)	70.21 (13.12)	16.97 (15.65)	(+24%)		
Student–teacher ratios	27.33 (0.23)	24.75 (0.36)	2.58*** (0.51)	(+10%)	145	35
Transfers received	1,147.48 (48.92)	1,130.76 (25.60)	16.72 (60.20)	(+2%)	334	35
Administrative indicators						
Correctness standards	79.30 (1.46)	71.40 (4.93)	7.90* (3.88)	(+11%)	236	35
Capacity	72.46 (1.56)	43.85 (4.20)	28.61*** (3.89)	(+65%)	236	35
Corruption	98.70 (6.03)	109.58 (22.94)	−10.88 (28.26)	(−10%)	89	32

Notes: Financial data are expressed in 2008 Colombian Pesos per pupil. Standard deviations of the means and standard errors of the difference within parentheses.

* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

responsibility and accountability brought about by decentralization, may have spurred higher education spending and led to an improvement in quality.[31] This would be in line with the conclusions by Faguet and Sanchez (2014), who find that in Colombia, more decentralization has led to more financial contributions of local governments, and that these are able to yield large improvements in public services. However, using the available data, I am able to give even stronger support to the local skills channel: I pin down large and significant differences in administration quality between highly developed and low-developed cities, in the direction one would expect. Low-developed cities appear to lack actual managerial capabilities rather than honest behavior. The level of skills among government officials has been revealed as a key driver for the success or failure of decentralized development projects in several countries, such as India, Pakistan, and Thailand, and to have heavily influenced local decentralization outcomes on a large scale in Kenya, Sudan, and Tanzania back in the 1970s and 80s (Rondinelli et al., 1983).

In summary – while bearing in mind that the analysis illustrated in this section does not claim to prove causal relationships and, instead, provides merely suggestive evidence – the data at hand, combined with the decentralization theory and results of past empirical studies, support the hypothesis that the large heterogeneity in local administration capacity may have played an important role in explaining the discrepant impacts that educational autonomy has brought to Colombian cities. An additional contribution by differences in local financial resources cannot be ruled out.

8 Conclusion

In this paper, I took advantage of an unusually simple administrative decentralization reform setting to demonstrate that cities characterized by different levels of local development have

31 I am abstracting from issues such as different funding *needs* between the two groups. For a discussion on why low-developed municipalities should however, if anything, have seen improvements in their financial situation after the reform, see Section A3 in the Appendix.

been affected differently by receiving managerial autonomy over the public education service. In the 10 years following the responsibility takeover, cities in the higher development deciles have significantly improved their local test score performance, while test scores have declined for the lowest-developed cities. The secondary school test performance I use as an outcome in this analysis is considered to be a good indicator of education quality in the country, and I find that the effect is not driven by apparent changes in the composition of test takers. The reform effects, and thus the education quality gaps, arise in the postreform periods and are decisively growing stronger over time. The empirical identification strategy I use – a municipal and year fixed effects model – is simple but robust and well suited to the context being analyzed. I am able to convincingly support the hypothesis that autonomous and nonautonomous municipalities were on similar performance trends before decentralization was implemented, even when allowing for different local development levels. Considering the characteristics of the Colombian reform, which left unchanged local taxation powers and the political structure, the treatment effects I estimate can be attributed to the sole devolution of managerial responsibility to the local level.

This paper represents a relevant contribution to the empirical literature on administrative decentralization outcomes, which often struggles to isolate well-identified treatment effects due to complexities in reform content or in the institutional context. A further desirable addition to the decentralization debate is given by the specific focus on local development levels as drivers of inequality, when combined with managerial autonomy.

Looking for clues on the channels of impact heterogeneity, I study financial and administrative postreform data for individual municipalities. Due to data limitations, I am not able to provide causal evidence on these channels; however, the analysis is helpful toward building informed hypotheses. Highly developed and low-developed municipalities appear to benefit of the same per-pupil transfers from the central government, but I am not able to exclude with certainty that highly developed municipalities have been devoting additional own resources to the financing of local public education. Importantly, highly developed municipalities perform significantly better on different local administration quality indicators with respect to low-developed cities, in particular, the indicators measuring skills and capacity at the local authority's disposal. One can picture the administrative bodies of low-developed cities being overwhelmed by the batch of new tasks and responsibilities that came along with the freedom over education management – and mistakes, delays, and bad decisions taking their toll on service quality. On the contrary, highly developed city administrations may have been better prepared to cope with the new duties, so that the benefits of decentralization dominated. These are conjectures formed on the basis of the data at hand and guided by the results of previous decentralization literature. They are consistent with this study's main finding that the devolution of autonomy on public education has caused remarkably heterogeneous results, depending on municipal development levels, so that highly developed and low-developed cities have been drifting apart in their performance over the postreform years.

Colombia has been one of the Latin American leaders in education investment over the past decades, and it distinguished itself among the countries with the fastest progress in education quality (Barrera- Osorio et al., 2012). Nevertheless, not all Colombian education reforms have achieved their desired results, and some have failed to safeguard equity in the distribution of impacts. The findings of this paper sound a note of caution in the design of decentralization

reforms, especially in contexts characterized by subnational heterogeneity in wealth and development. Oversimplified decentralization criteria might prove useful for subsequent impact evaluation studies but are certainly not the best guarantee of equity in outcomes across the involved parties. When handing responsibilities in public service delivery to the local level, less-advantaged local authorities may need additional and well-planned support in order to prevent regional inequality from growing and decentralization backfiring for some segments of the population.

Declarations

Availability of data and material

Not applicable.

Competing interests

The author declares that she has no competing interests.

Funding

The author is grateful for financial support received during the course of this project from the Luca d'Agliano Foundation (Turin, Italy).

Author contributions

Not applicable.

Acknowledgments

The author is grateful for helpful comments and suggestions by Jérôme Adda, Jesús Arias Duarte, Oriana Bandiera, Caroline Hoxby, Matthias Hübener, Andrea Ichino, Alan Manning, Éric Maurin, Guy Michaels, Alexander Moore, Steve Pischke, Fabio Sánchez Torres, Olmo Silva, and a number of anonymous referees. Gratitude for help with data goes to Germán Cano Torres at DNP. Many thanks are also due to seminar participants at the London School of Economics, the European University Institute, the University of Los Andes, and several conferences.

References

Alesina, A.; A. Carrasquilla; J. J. Echavarría (2000): Decentralization in Colombia, In: Fedesar- rollo Working Paper Series 15.

Angrist, J.; V. Lavy (1999): Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. The Quarterly Journal of Economics 114(2), 533-575.

Angrist, J.; J.-S. Pischke (2009): Mostly Harmless Econometrics: an Empiricists Guide. Princeton: Princeton University Press.

Autor, D. H. (2003): Outsourcing at will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. Journal of Labor Economics 21(1), 1-42.

Bardhan, P.; D. Mookherjee (2000): Capture and Governance at Local and National Level. American Economic Review [PaP], 90(2), 135-139.

Bardhan, P.; D. Mookherjee (2002): Decentralization of Governance and Development. Journal of Economic Perspectives 16(4), 185-205.

Bardhan, P.; D. Mookherjee (2005): Decentralizing Antipoverty Program Delivery in Developing Countries. Journal of Public Economics 89(4), 675-704.

Bardhan, P.; D. Mookherjee (2006): Decentralisation and Accountability in Infrastructure Delivery in Developing Countries. The Economic Journal 116(508), 101-127.

Barrera-Osorio, F.; D. Maldonado; C. Rodríguez (2012): Calidad de la educación básica y Media en Colombia: Diagnóstico y propuestas. Serie Documentos de Trabajo Universidad del Rosario 126.

Besley, T.; R. Burgess (2002): The Political Economy of Government Responsiveness: Evidence from India. The Quarterly Journal of Economics 117(4), 1415-1451.

Besley, T.; R. Burgess (2004): Can Labor Regulation Hinder Economic Performance? Evidence from India. The Quarterly Journal of Economics 119(1), 91-134.

Blair, H. (2000): Participation and Accountability at the Periphery: Democratic Local Governance in Six Countries. World Development 28(1), 21-39.

Bonet, J.; G. J. Pérez V.; J. Ayala (2014): Contexto histórico y evolución del SGP en Colombia. Banco de la República - CEER Cartagena.

Borjas, G. J.; O. L. Acosta (2000): Education Reform in Colombia. Bogota: FEDESARRCLO. Working Papers Series 19.

Breton, A.; A. Scott (1978): The Economic Constitution of Federal States. Toronto: University of Toronto Press.

Caballero, P., ed. (2006): Foro "Descentralización y certificación en educación, experiencias y de- safíos, el caso colombiano". Bogotá, D.C.: PREAL GDyÁ - Conversemos sobre educación.

Cerquera, D.; P. Jaramillo; N. Salazar (2000): La educación en Colombia: evolución y diagnostico. Bogotá D.C.: Departamento Nacional de Planeación.

Congreso. (1993): *Ley 60 de 1993 - Por la cual se dictan normas orgánicas sobre la distribución de competencias de conformidad con los artículos 151 y 288 de la Constitución Política y se distribuyen recursos según los artículos 356 y 357 de la Constitución Política y se dictan otras disposiciones.* Colombia: Congreso de la República.

Congreso. (1994): Ley 115 de 1994 - Por la cual se expide la ley general de educación. Colombia: Congreso de la República.

Congreso. (2001): Ley 715 de 2001 - Por la cual se dictan normas orgánicas en materia de recursos y competencias de conformidad con los artículos 151, 288, 356 y 357 (Acto Legislativo 01 de 2001) de la Constitución Política y se dictan otras disposiciones para organizar la prestación de los servicios de educación y salud, entre otros. Colombia: Congreso de la República.

Corte Constitucional, C. (1997): Sentencia SU.559/97. Corte Constitucional de Colombia.

Cortés, D. (2010). Do More Decentralized Local Governments do Better? An Evaluation of the 2001 Decentralization Reform in Colombia. Serie documentos de trabajo - Universidad del Rosario 84.

Cronin, J.; G. G. Kingsbury; M. S. McCall; B. Bowe (2005): The Impact of the No Child Left Behind Act on Student Achievement and Growth: 2005 Edition. Non-Journal ED491250. OR: Northwest Evaluation Association, p. 75.

Crook, R.; J. Manor (1998): Democracy and Decentralisation in South Asia and West Africa: Participation, Accountability and Performance. Cambridge: Cambridge University Press.

DDTS, D. (2004): Evaluacion del sistema general del participaciones 2003. Bogotá, D.C.: Departamento Nacional de Planeación - Dirección de Desarrollo Territorial.

Directiva Ministerial (2003): In collab. with MEN.

DNP (2002): Documento Conpes Social n.57. Bogotá, D.C.: Departamento Nacional de Planeación.

DNP (2004a). Documento Conpes Social n.77. Bogotá, D.C.: Departamento Nacional de Planeación.

DNP (2004b). Documento Conpes Social n.89. Bogotá, D.C.: Departamento Nacional de Planeación.

DNP (2011): Evaluación del desempeño integral de los municipios. Bogotá D.C.: Departamento Nacional de Planeación - Dirección de Desarrollo Territorial.

DNP (2014): Orientaciones para realizar la evaluación del desempeño integral municipal, Vigencia 2013. Bogotá, D. C.: Departamento Nacional de Planeación - Dirección de Desarrollo Territorial.

Faguet, J. P. (2004): Does Decentralization Increase Responsiveness to Local Needs? Evidence from Bolivia. Journal of Public Economics 88, 867-894.

Faguet, J. P.; F. Sanchez (2008): Decentralization's Effects on Educational Outcomes in Bolivia and Colombia. World Development 36(7), 1294-1316.

Faguet, J. P.; F. Sanchez (2014): Decentralization and Access to Social Services in Colombia. Public Choice 160(1), 227-249.

Fernández Ariza, M.; A. García Morales (1999): Inequidad interregional en el situado fiscal para educación. Investigación y Desarrollo. Universidad del Norte. 9, 29-46.

Ferraz, C.; F. Finan (2011): Electoral Accountability and Corruption: Evidence from the Audits of Local Governments. American Economic Review 101(4), 1274-1311.

Figlio, D.; S. Loeb (2011): Chapter 8 - School Accountability, in: Handbook of the Economics of Education. Hanushek, E. A.; S. J. Machin; L. Woessmann (eds.), North Holland (Amsterdam), Vol. 3, 383-421.

Galiani, S.; P. Gertler; E. Schargrodsky (2008). School Decentralization: Helping the Good Get Better, but Leaving the Poor Behind. Journal of Public Economics 92(10), 2106-2120. ISSN: 0047-2727. doi:10.1016/j.jpubeco.2008.05.004.

Gómez, A. S.; L. P. Tovar; C. Alam (2001): Situación de la educación básica, media y superior en Colombia. Casa Editorial El Tiempo - Fundación Corona - Fundación Antonio Restrepo Barco.

Granger, C. W. (1969): Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. Econometrica: Journal of the Econometric Society 37, 424-438.

Hammond, G. W.; M. S. Tosun (2011): The Impact of Local Decentralization on Economic Growth: Evidence from U.S. Counties. Journal of Regional Science 51(1), 47-64.

ICFES (2013): Sistema Nacional de Evaluación Estandarizada de la Educación - Alineación del examen SABER 11. Bogotá, D.C.

Iregui, A. M.; L. Melo; J. E. Ramos (2006): Evaluación y análisis de eficencia de la educación en Colombia. Bogotá, D.C.: Banco de la Republica, Colombia.

Juetting, J.; E. Corsi; C. Kauffmann; I. McDonnell; H. Osterreider; N. Pinaud ; L. Wegner (2005): What Makes Decentralization in Developing Countries Pro-Poor? The European Journal of Development Research 17(4), 626-648.

Kaiser, K. (2006): Decentralization Reforms. Analyzing the Distributional Impact of Reforms. Ed. by A. Coudouel and S. Paternostro. Vol. 2. Washington: World Bank.

Lonzano, L. I., J. E. Ramos; H. Rincón (2007): Implicaciones Fiscales y Sectoriales de la Reforma a las Transferencias Territoriales en Colombia. Borradores de Economía - Banco de la Republica, Colombia 437.

MacKinnon, J.; M. Webb (2017): Wild Bootstrap Inference for Wildly Different Cluster Sizes. Journal of Applied Econometrics 32(2), 233-254.

MacKinnon, J.; M. Webb (2020, forthcoming): Randomization Inference for Difference-in-Differences with Few Treated Clusters. Journal of Econometrics.

Manor, J. (1999): The Political Economy of Democratic Decentralization. Washington: World Bank.

McCrary, J. (2008): Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. Journal of Econometrics. The Regression Discontinuity Design: Theory and Applications 142(2), 698-714. ISSN: 0304-4076. doi:10.1016/j.jeconom.2007.05.005.

Musgrave, R. A. (1959): The Theory of Public Finance. New York: McGraw-Hill.

Oates, W. (1972): Fiscal Federalism. New York: Harcourt Brace Jovanovich.

OTyA (2017): Indicador de Sanciones Penales. Observatorio Transparencia y Anticorrupción - Fuente: Fiscalía General de la Nación.

Panizza, U. (2004): What Drives Fiscal Decentralisation? CESifo DICE Report 1/2004.

Reinikka, R.; J. Svensson (2004): Local Capture: Evidence from a Central Government Transfer Program in Uganda. The Quarterly Journal of Economics 119(2), 678-704.

Rondinelli, D. A. (1981): Government Decentralization in Comparative Perspective - Theory and Practice in Developing Countries. International Review of Administrative Sciences 47(2), 133-145.

Rondinelli, D. A.; J. R. Nellis; G. S. Cheema (1983): Decentralization in Developing Countries: A Review of Recent Experience. World Bank Staff Working Papers - Management and Development Series 581.8.

Roodman, D.; J. MacKinnon; M. Nielsen; M. Webb (2019): Fast and Wild: Bootstrap Inference in Stata Using Boottest. Stata Journal 19(1), 4-60.

Santa Maria Salamanca, M.; P. Camacho Alvarez; J. F. Arias Duarte (2001): Exposición de Motivos de la Reforma a la Ley 60 de 1993. Sector Educación y Sector Salud. 173. Bogotá D.C.: DNP - Dirección de Desarrollo Social.

Santa Maria Salamanca, M.; N. Millan; J. Moreno; C. F. Reyes (2009): La descentralización y el financiamento de la salud y la educación en los departamentos: ¿Cuáles son las alternativas? Informe final. Federación nacional de departamentos y Fedesarrollo.

Sarmiento, A.; J. E. Vargas (1997): Descentralización de los servicios de educación y salud en Colombia. Conyuntura Social - Fedesarrollo/Instituto SER de Investigación 16, 91-136.

Toro, G. (2006): Trabajamos por la descentralizacion, la gobernabilidad y la autonomia local. Pre-sentacion. Federación Colombiana de Municipios (FCM).

Vergara, C. H.; M. Simpson (2001): Evaluación de la descentralización municipal en Colombia. Estudio general sobre los antecedentes, diseño, avances y resultados generales del proceso de descentralización territorial en el Sector educativo. Archivos de Economía - Departamento Nacional de Planeación 168.

Zhang, X. (2006): Fiscal Decentralization and Political Centralization in China: Implications for Growth and Inequality. Journal of Comparative Economics 34(4), 713-726.

Appendix

A.1 Descriptive Statistics

Table A1 Descriptive statistics

	All cities	Autonomous	Nonautonomous
Test-related			
Test scores (pre-reform)	41.44 (1.52)	41.74 (1.34)	41.42 (1.53)
Test scores (postreform)	42.51 (2.30)	43.39 (2.35)	42.46 (2.29)
Number of test takers	296.16 (483.57)	1,839.40 (1,116.16)	213.70 (214.11)
Share of low-SES test takers	0.88 (0.10)	0.78 (0.11)	0.87 (0.10)
Municipal characteristics			
Municipal Development Index 2001	36.03 (10.81)	50.67 (10.25)	35.25 (10.27)
Population in 2002 (thousands)	36,834.28 (52,225.38)	221,149.10 (106,137.50)	26,855.00 (18,259.58)
Education finance (postreform)			
Total spending (per student)	-	1,159.46 (398.03)	-
Salaries	-	925.72 (297.42)	-
Infrastructure and materials	-	99.63 (90.55)	-
Others	-	84.09 (93.91)	-
Student–teacher ratio	-	26.85 (2.59)	-
Transfers received (per student)	-	1,144.58 (416.21)	-
Number of cities	692	35	658

Notes: Variable means and (standard deviations). Data on city-level education finance is only available for autonomous cities. Financial data is in 2008 Colombian pesos.

SES, socioeconomic status.

A.2 Additional Tables and Figures

Figure A1 Maps of Colombia: a) Colombia's departments; b) Colombia's municipalities in black, those which were assigned autonomy over education in 2002; c) Distribution of Municipal Development Index in 2001. *Note:* In maps b) and c), the rural southeast is omitted to increase readability in the densely populated area.

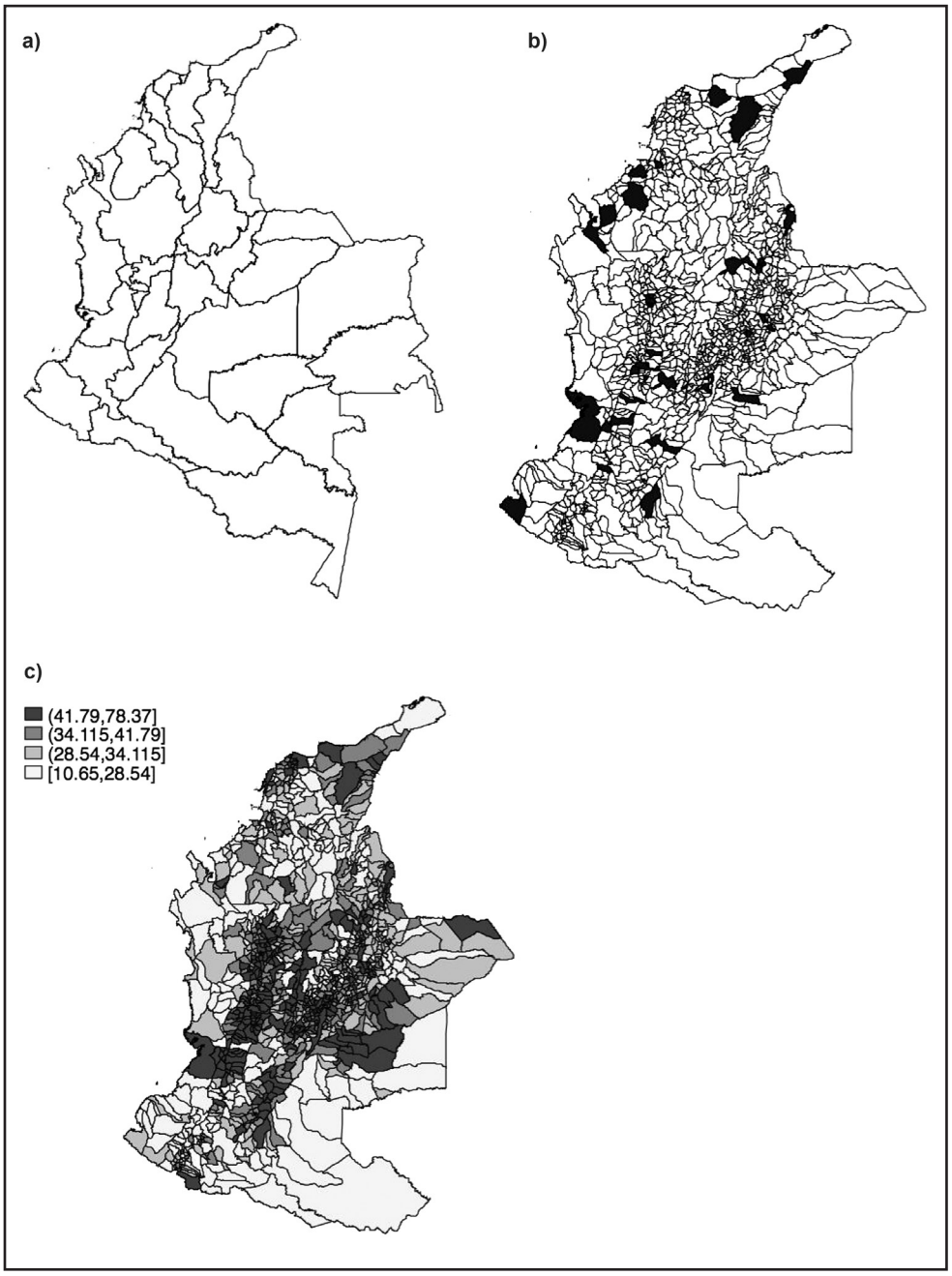


Figure A2 Municipal development index in the empirical sample. *Notes:* The figure plots the 2001 MDI values for our sample of cities. The normal curve is overlaid.

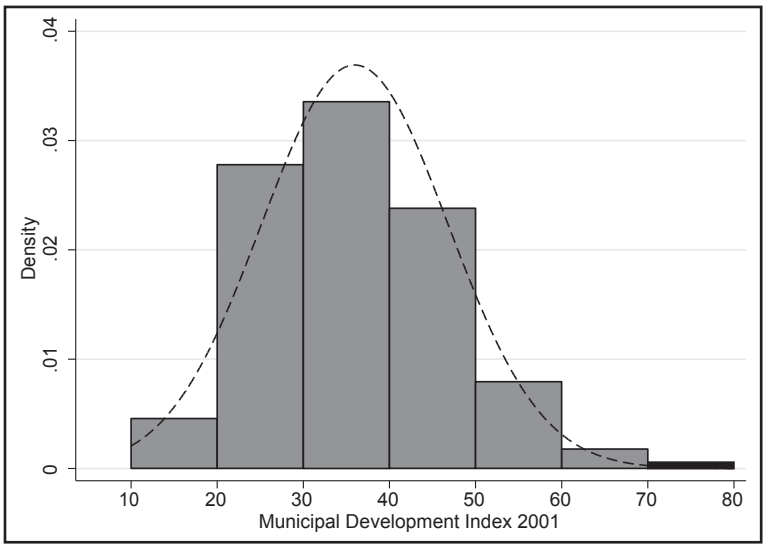


Figure A3 Evolution of test scores for the two groups of cities over time. *Notes:* The figure shows the evolution of test score averages over time. Autonomous cities are solid, nonautonomous ones are dashed; highly developed cities are darker, and low-developed ones are lighter. In subfigures (a) and (b) at the bottom, differences between the postreform gap and the pre-reform (2001) gap are indicated ($H0$: $diff = 0$). Pre-reform and transition years are shaded in gray.

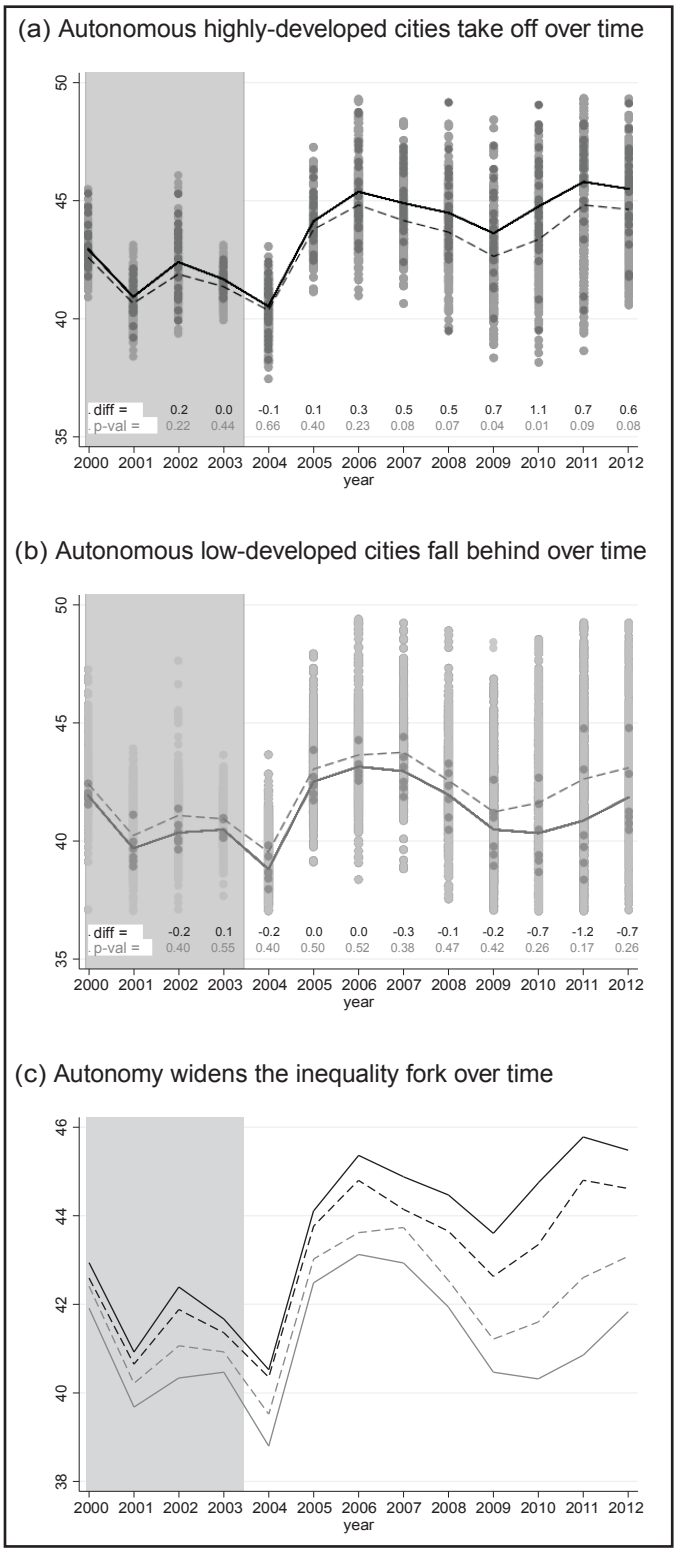


Table A2 The effect of municipal autonomy on test scores – other subjects

Subject	Period average		Over-time evolution				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Naive	Baseline	2002–2003	2004–2005	2006–2007	2008–2009	2010–2012
Critical reading							
Autonomy	0.28*** (0.11)	−0.56 (0.42)	−0.47* (0.27)	−0.01 (0.78)	0.33 (0.75)	0.10 (0.59)	−2.13*** (0.66)
Autonomy × MDI		0.02** (0.01)	0.01 (0.01)	0.00 (0.01)	0.00 (0.02)	0.00 (0.01)	0.06*** (0.01)
Sciences							
Autonomy	0.21*** (0.08)	−0.58* (0.32)	−0.65 (0.52)	−0.13 (0.38)	−0.06 (0.41)	−0.08 (0.39)	−1.65*** (0.55)
Autonomy × MDI		0.02** (0.01)	0.02* (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.04*** (0.01)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	8,734	8,734	2,742	2,665	2,707	2,705	3,367
N groups	692	692	692	692	692	692	692
R-sq.	0.74	0.74	0.46	0.75	0.65	0.40	0.65

Notes: SEs clustered by municipalities in parentheses.

SE, Standard Errors.

* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A3 The effect of municipal autonomy on number of test takers (% change)

	Period average		Over-time evolution				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	naïve	Baseline	2002–2003	2004–2005	2006–2007	2008–2009	2010–2012
Autonomy	−0.16*** (0.04)	−0.03 (0.20)	−0.08 (0.12)	−0.19 (0.48)	0.11 (0.20)	−0.16 (0.26)	0.13 (0.21)
Autonomy × MDI		0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	−0.01* (0.00)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	8,734	8,734	2,742	2,665	2,707	2,705	3,367
N groups	692	692	692	692	692	692	692
R-sq.	0.27	0.27	0.05	0.10	0.35	0.09	0.57

Notes: SEs clustered by municipalities in parentheses.

SE, Standard Errors.

* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A4 The effect of municipal autonomy on share of low-SES test takers

	Period average		Over-time evolution			
	(1)	(2)	(3)	(4)	(5)	(6)
	naïve	Baseline	2002–2003	(2004–2007)	2008–2009	2010–2012
Autonomy	0.01 (0.01)	0.06 (0.04)	0.02 (0.08)		0.12** (0.05)	0.04 (0.07)
Autonomy × MDI		0.00 (0.00)	0.00 (0.00)		0.00** (0.00)	0.00 (0.00)
Time dummies	Yes	Yes	Yes		Yes	Yes
N	6,088	6,088	2,742		2,705	3,367
N groups	692	692	692		692	692
R-sq.	0.19	0.19	0.01		0.13	0.34

Notes: SEs clustered by municipalities in parentheses. In the period 2004–2007, socioeconomic status (SES) data were not collected.

SE, Standard Errors.

* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

A.3 Change in Financial Transfers

As described in Section 3.2.1, the 2001 decentralization reform brought along a change in the formulas determining central government transfers to be used by local governments for the delivery of public services. The main differences can be summarized as follows: 1) elimination of the volatility in amounts transferred, which were previously determined as a percentage of national fiscal revenues and thus subject to the economic cycle; and 2) a shift in the transfer allocation criteria away from a cost-of-service perspective and toward a number-of-users perspective. Overall, in the years following the reform, real transfers to local authorities exhibited a stable upward trend across all service sectors (Bonet et al., 2014).

Again, the main interest lies in discussing whether the change in transfer formulas might play a role in explaining the estimation results, especially the heterogeneity in impact across the local development spectrum. Considering that the changes in funding applied to all local authorities in the nation, they will be absorbed by the time fixed effects embedded in the baseline estimation model (2). Similarly, results would not be affected if funding changed asymmetrically for *all* highly developed or *all* low-developed cities. The only potentially worrying case is that involving alterations that are *asymmetric across development and autonomy status*, where the possible alternative cases of asymmetry are in their substance very similar to those listed in subsection 6.3 about the "reverse treatment" conjecture.

Assuming that higher funding helps in improving test scores, one would be worried about the case in which the change in funding had favored more-developed cities, *especially so if autonomous*, and had disadvantaged lower-developed cities, *especially so if autonomous*. In this situation, the test-score divergence I find might potentially be in part or entirely explained by changes in funding rather than by managerial autonomy.

Due to the lack of data on the amount of funding effectively reaching nonautonomous municipalities under departmental supervision, it is not possible to rule out the above-mentioned possibility empirically. It nevertheless seems reasonable to draw the following observations:

(1) The transition to a transfer system giving more weight to pupil counts should have, if anything, favored municipalities characterized by *low* levels of local development. Such areas had historically been financially disadvantaged because of proportionally fewer education staff on their payrolls, subsequent deflated service costs, and ensuing lower transfers (Corte Constitucional, 1997, par.19; Ariza and Morales, 1999; Cerquera et al., 2000; Salamanca et al., 2001).
(2) Autonomy eliminates the departmental "filter" between financial resource entitlement and effective financial resource reception. Therefore, if anything, autonomous low-developed cities should have benefited from the change in formulas in a somewhat quicker and fuller way with respect to their nonautonomous counterparts.

The combination of observations 1) and 2) would appear to discredit the scenario which would point at changes in funding as explaining our education performance results to any significant extent.

A.4 Robustness

A.4.1 Inclusion of controls

In Table A5 in Appendix, I show the results of augmenting Model 2 with different municipal-level control variables.[32] In Column (1), I add municipal population in 10,000 inhabitants. In Column (2), I add two variables proxying for the quality of municipal health services – a test that may be of particular interest given that these were decentralized at the same time as education (see Section 3.2). The first variable indicates the share of population affiliated to public health care, and the second variable indicates the share of newborns born with a low birth weight. In Column (3), I also add a rurality index, measuring the share of population living in rural areas, which serves the purpose of controlling for potential over-time changes in population distribution and living conditions. In the same spirit, Column (4) adds two inequality indexes: a classical Gini index and a Gini index specific to land ownership.

The inclusion of these controls does not affect the estimation of the impact of education autonomy or of its development gradient, backing the validity of the identification assumptions.

32 *Source*: "*Indicadores para los ODS*" - *Centro de datos CEDE - Facultad de Economia, Universidad de Los Andes*. Retrieved in June 2019 at https://datosods.uniandes.edu.co/.

Table A5 The effect of municipal autonomy on test scores - with controls

	Controls				
	(1)	(2)	(3)	(4)	(5)
	Population	+ **health**	+ **rurality**	+ **inequality**	**Baseline**
Autonomy	−2.68*** (0.93)	−2.45*** (0.91)	−2.69*** (0.90)	−2.65** (1.15)	−2.53*** (0.87)
Autonomy × MDI	0.06*** (0.02)	0.05*** (0.02)	0.06*** (0.02)	0.06*** (0.02)	0.06*** (0.02)
Population (10k)	0.17*** (0.05)	0.16*** (0.05)	0.18*** (0.05)	0.18*** (0.06)	
Health (registered)		−0.54 (0.45)	−0.55 (0.46)	−0.49 (0.43)	
Health (low BW)		−164.08*** (59.28)	−178.02*** (59.27)	−175.47*** (61.71)	
Rurality index			−5.19*** (1.60)	−4.05** (1.71)	
Gini index				1.52*** (0.49)	
Gini index (land)				−0.50 (1.12)	
Municipality FE	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes
N	8,734	8,716	8,716	7,910	8,734
N groups	692	692	692	692	692
R-sq.	0.51	0.52	0.52	0.51	0.51
Mean *y*	42.34	42.34	42.34	42.34	42.34

Notes: SEs clustered by municipalities in parentheses.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

A.4.2 Discontinuity samples

Even though it excludes the very smallest and very largest cities, the main sample of analysis encompasses communities whose sizes are quite heterogeneous. One potential concern with this setup is that with city sizes diverging significantly, the number and nature of confounders affecting the analysis may grow. In this robustness check, I limit the sample to municipalities closer to the autonomy threshold, both from the left and from the right. Following Angrist and Lavy (1999), I refer to these as "discontinuity samples". The baseline specification (2) is now estimated on samples of cities that are similar to each other in size, of which some cities acquired education autonomy in 2002 and some did not.

Table A6 Results on "discontinuity samples" around the autonomy threshold

	Discontinuity samples				
	(1)	(2)	(3)	(4)	(5)
	± 20,000	± 30,000	± 40,000	± 50,000	Baseline
Autonomy	−6.25*** (0.58)	−3.15** (1.25)	−2.65** (1.18)	−2.45** (1.17)	−2.53*** (0.87)
Autonomy × MDI	0.15*** (0.02)	0.07** (0.03)	0.06** (0.03)	0.06** (0.03)	0.06*** (0.02)
Municipality FE	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes
N	294	473	740	1,086	8,734
N groups	23	37	58	85	692
R-sq.	0.72	0.70	0.66	0.63	0.51
Mean *y*	43.00	42.79	42.80	42.60	42.34

Notes: SEs clustered by municipalities in parentheses.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A6 shows the results of the exercise repeated with samples including cities within 20, 30, 40, and 50 thousand inhabitants above and below the autonomy threshold. The last column in the table recalls our baseline results for comparison. As discontinuity samples become tighter around the threshold, results are augmented – in the sense that the linkage between local development and the impact of autonomy on local education quality is stronger than in the baseline. As the boundaries around the threshold expand to 40,000 inhabitants and more, results seem to stabilize around the baseline values. From this exercise, one may conclude that confounders related to city size do not seem to play an important role in explaining the main message of this paper. If anything, when moving very close to the autonomy threshold, the key message strengthens even further.

A.4.3 Time trends and different standard errors

Table A7 in Appendix shows the results of the second exercise described in Section 6.1. In columns (1)–(4), different kinds of development-specific time trends are added to the baseline model (2) in order to test the robustness of the treatment effect estimates.

Column (5) implements a different robustness check: standard errors are clustered at the department level, instead of at the municipal level. This is done in order to account for the possibility that errors be correlated within a department, for instance, because of several municipalities being managed by the same department. In this specification, standard errors remain virtually unchanged, and the statistical significance of the main results is unaffected.

Inference derived applying traditional nonparametric bootstrapping and the more recent wild-clustered bootstrapping[33] techniques (MacKinnon and Webb, 2017, 2020, forthcoming; Roodman et al., 2019) yields somewhat larger confidence intervals around the results, but without affecting their robustness in any significant way (available on request).

33 Thanks to an anonymous referee for suggesting this additional check, motivated by the possibility that traditional cluster-robust SEs might suffer from asymptotic bias due to the fact that there is a small number of treated clusters and that cluster sizes differ within the sample.

Table A7 Development-specific time trends: department-level SEs

	IDM-specific trends				Department-level SEs	
	(1)	(2)	(3)	(4)	(5)	(6)
	</> 40	Tertiles	Quartiles	Quintiles	Baseline	Baseline
Autonomy	−1.58* (0.87)	−1.71** (0.75)	−1.66* (0.88)	−1.67** (0.81)	−2.53*** (0.89)	−2.53*** (0.87)
Autonomy × MDI	0.04** (0.02)	0.04*** (0.01)	0.04** (0.02)	0.04** (0.02)	0.06*** (0.02)	0.06*** (0.02)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
IDM-specific *T* trends	Yes	Yes	Yes	Yes	No	No
N	8,734	8,734	8,734	8,734	8,734	8,734
N groups	692	692	692	692	692	692
R-sq.	0.53	0.53	0.54	0.53	0.51	0.51
Mean *y*	42.34	42.34	42.34	42.34	42.34	42.34

Notes: Columns (1)–(4) and (6): SEs clustered by municipalities in parentheses. Column (5): SEs clustered by departments in parentheses.

IDM, *Índice de Desarrollo Municipal.*

* $p<0.10$, ** $p<0.05$, *** $p<0.01$.